Muscle strength in adults with spinal cord injury: a systematic review protocol of manual muscle testing, isokinetic and hand held dynamometry clinimetrics

Kerry Peek BSc (Hons), PG Cert¹
Prapaphan Pensuk RN, BNS, MNS²

1. Joanna Briggs Institute, University of Adelaide,
2. Lyell McEwin Hospital, South Australia,

Corresponding author
Kerry Peek
kerry.peek@adelaide.edu.au

Review question/objective

The primary objective of this systematic review is to examine the best available evidence on the clinimetrics of the three main methods of assessing muscle strength: manual muscle testing (MMT), isokinetic dynamometry (ID) and hand held dynamometry (HHD). The clinimetric domains include:

• Reliability
• Validity
• Responsiveness
• Interpretability.

A further objective is to determine whether it is appropriate for HHD to become the standard tool for measuring muscle strength for adults with spinal cord injury.

Background

Spinal cord injury is often defined as damage either through trauma or disease to the spinal cord which results in a loss of function such as motor control and sensation. If the injury to the spinal cord occurs in the cervical spine, the resultant loss of function will involve the upper limbs, trunk and lower limbs, which is termed tetraplegia. The term paraplegia relates to functional loss which involves the trunk and lower limbs only. The American Spinal Injury Association (ASIA) further defines spinal cord injury as complete or incomplete, using the ASIA Impairment Scale. According to this scale, an incomplete spinal cord injury presents with no preservation of motor or sensory function below the sacral segments of S4-5 (in other words no rectal sensation or voluntary anal sphincter control). Whereas an incomplete spinal cord injury has preserved sensory function below the spinal level of injury including S4-5, with or without motor function. In Australia, the spinal cord injury prevalence rate among individuals 15 years and older has been reported as 681 per million population.¹ This rate is

doi: 10.11124/jbisrir-2013-969
considered lower than that in the United States but higher than that in a number of European countries. With increasing survival and little variation in the incident case number evident from data collected since 1986, the prevalent population of spinal cord injuries in Australia is increasing.\textsuperscript{1}

Spinal cord injury is arguably one of the most devastating medical conditions. It can cause considerable life changing consequences in all facets of function and existence\textsuperscript{2} and requires extensive rehabilitation.

Physiotherapists and many other health professionals routinely conduct muscle strength tests of their patients as part of the objective examination. Following spinal cord injury, the determination of muscle strength is important for a number of reasons including neurological classification of injury, therapeutic planning and outcome evaluation.\textsuperscript{3}

Accurate baseline assessment of muscle strength in the initial examination of an adult with spinal cord injury is important for physiotherapists as it will direct them in their treatment plan and the treatment goal-setting process. Spinal cord injuries and their subsequent impairment tend to vary between individuals, particularly in relation to the number of incomplete lesions. With an incomplete lesion the individual may have varying muscle activity below the main spinal level of injury which needs to be carefully assessed and measured so that no muscle activity is missed which could have a substantial impact on the individual's treatment and exercise focus, and therefore functional recovery. Accurate muscle strength testing is necessary to highlight areas of muscle weakness or indeed strength as well as muscle imbalances between agonistic and antagonistic muscle groups in order to direct the exercise component of the rehabilitation program.

Accurate retesting of muscle strength is also required at follow up appointments as an indicator of physiotherapeutic intervention efficacy. The progressive evaluation of physiotherapy treatment outcomes is an integral part of professional accountability and is a requirement of the Australian Physiotherapy Competency Standards.\textsuperscript{4} There is an emphasis on clinical outcome measures for physiotherapists during rehabilitation of their patients, and accurate measurement of muscle strength can provide quantifiable data in support of patient outcomes. A significant percentage of traumatic spinal cord injuries are caused through road traffic accidents. Funding for physiotherapy through external agencies such as the Motor Accident Commission requires the physiotherapist to provide evidence of improvement and to state the outcome measures used in support of the need for ongoing physiotherapy. Muscle strength data can assist with this.

A number of methods are available for assessing muscle strength. MMT commonly takes the form of requesting the patient to perform a muscle activity which is then graded by its ability to act in a gravity eliminated position, its ability to act against gravity or by resistance applied by the therapist.\textsuperscript{3} The therapist may then record this subjectively utilizing their experience as to whether they felt a muscle or action was impaired or not. They may quantify this muscle activity using the scale developed by the Medical Research Council (MRC), which uses an ordinal scale of 0-5. However, this scale also relies somewhat on the therapist's experience. In addition, the use of the MRC scale may be considered valid within an individual but not across individuals as it is not safe to assume that one person's grade 4 is the same as another person's. Just as grade 4 is greater than grade 2 but it does not equal twice as much strength as grade 2, as the interval between grades is not equal on the MRC scale.

As early as 1916 the dissatisfaction with imprecise and subjective measures of muscle strength was reported.\textsuperscript{5} By 1956, evidence in the use of HHDs to objectify muscle strength tests was being demonstrated.\textsuperscript{6} HHDs are portable, inexpensive devices which can be used to obtain more objective measures of muscle strength.\textsuperscript{7}
Further developments in the use of dynamometry over the last 35 years or so have seen the emergence of computerized testing devices known as IDs, which was reported as the 'gold standard' of measuring muscle strength in a systematic review by Stark et al. in 2011. The main differences between the HHD and the ID are ease of use, cost (an ID can cost 40 times more than a HHD), size and access.

MMT in contrast requires no equipment and is therefore the most widely-used clinical method.

However, where MMT uses ordinal data, dynamometry uses the highest level of measurement, the ratio scale. As it uses a scale where the units of measure have equal intervals with true zero, comparisons can be made across individuals and this allows for more mathematical and statistical operations with the data.

The objective of this systematic review is to consider the clinimetrics for each of these methods of assessing muscle strength specific to spinal cord injury. The methodological quality of the studies to be included will be based on the COnsensus-based Standards for the selection of health Measurement INstruments (COSMIN) checklist. The aim of the COSMIN initiative is to improve the selection of health measurement instruments including the development of a COSMIN critical appraisal tool (checklist). This checklist was developed in an international Delphi study as a multidisciplinary, international collaboration. For the purpose of this review a number of terms has been defined as these will be the clinimetric domains:

Reliability: the degree to which the measurement of muscle strength is free from measurement error. This includes repeated measurement under several conditions such as over time (test-retest), by different people (inter-rater) or by the same person (intra-rater). Reliability will be considered for MMT, IDs and HHDs.

Validity: the degree to which the muscle strength test measures the construct it purports to measure. This includes face validity (the degree to which the muscle strength test adequately reflects a measurement of muscle strength) and criterion validity (the degree to which the measurements data is an adequate reflection of a 'gold standard').

Responsiveness: the ability of the muscle strength test to detect muscle strength changes over time.

Interpretability: the degree to which one can assign clinical meaning to the quantitative scores or change in scores in muscle strength.

After reviewing the best available evidence, is it appropriate for HHDs to become the standard tool for measuring muscle strength in adults with spinal cord injury?

The role of measurement in physiotherapy practice cannot be overstated. Physiotherapists are moving towards choosing assessment procedures that have been shown in research to be reliable and valid, and which increases the likelihood of accurately measuring the impact of an intervention or change over time.

Clinically it is important to understand the clinimetrics of the method being used to assess an individual's muscle strength. For example, a literature review by Sisto and Dyson-Hudson reported that instrumented strength measures produce greater reliability than MMT, generally because of their measurement quality. However, in the case of ID which may not be available to the clinician either due to access or the time taken to set up or calibrate the machine it may be more practical to use MMT or HHDs.
In the case of measurement reliability, the involvement of a specified protocol designed to maximize the reliability of the method for assessing muscle strength such as standardized testing procedures means measurement errors can be reduced.

When we consider validity, we may consider that a dynamometer is a valid test for measuring muscle strength because the transducer responds with a signal proportionate to the exerted force. However, when the validity of inferences is considered, we must also look at what the value or score means to the individual being tested and how the data can be interpreted clinically. We must also consider the responsiveness of the method being used. We know MMT uses ordinal data which may not be responsive to small changes in muscle strength whereas the ratio data from dynamometry may be more sensitive to change.

The American Spinal Injury Association (ASIA) first published a standard system for neurological assessment and classification of spinal cord injury in 1982 which was later adopted by the International Medical Society of Paraplegia (IMSOP).\textsuperscript{11} The ASIA classification scale uses MMT in the measurement of muscle strength. However, clinimetrically, is MMT the most appropriate assessment method to use or does research support a need for instrumented muscle strength assessment such as via the use of HHDs?

By reviewing the primary research into the clinimetrics of MMT, IDs and HHDs in the assessment of muscle strength, physiotherapists and other clinicians can be guided towards using the most reliable, valid, responsive and interpretable method for adults with spinal cord injury. A systematic review can therefore answer whether there is support in the research to recommend HHDs as the standard tool for measuring muscle strength for adults with spinal cord injury. In addition, gaps in primary research may also be highlighted. A search of the Joanna Briggs Library did not identify any systematic reviews focusing on the clinimetrics of MMT, IDs and HHDs in the assessment of muscle strength in individuals with spinal cord injury. This provides an opportunity for the systematic search, synthesis and summary of the best available evidence in this area.

**Keywords**

dynamometry; clinimetrics; manual muscle strength; isokinetic; hand held dynamometry; spinal cord injury

**Inclusion criteria**

**Types of participants**

Adult population (adult being defined as 15 years and older in reference to the Australian Spinal Injury Registry which categorises young adults as >15 years) with spinal cord injury, regardless of cause, time since injury, gender or ethnicity.

**Tests**

All studies evaluating MMT, IDs and HHDs in all of their variations administered by any health professional.

**Types of outcomes**

The objective outcomes will be measured in terms of the clinimetric domains of reliability, validity, responsiveness and interpretability for MMT, IDs and HHDs, as defined above.
Types of studies

All quantitative clinimetric studies regardless of design will be sought. Studies which report on the clinimetrics of MMT, IDs and/or HHDs, as well as those using some form of comparison between the methods used will be considered.

Search strategy

A three-step search strategy will be utilised for this systematic review. The initial search will include Cochrane database, Pubmed, CINAHL, SPORTSDiscus, OT seeker, PEDro as well as the specific databases for Australian Physiotherapy Association Journal and The Chartered Society of Physiotherapy Journal as well as Spine.

The initial search will be proceeded by an analysis of the text words contained within the title and abstract, and of the index terms used to describe the study.

A second search using all identified keywords and indexed terms will then be commenced across all included databases.

The third step will look at the reference list of all of the above for any additional studies. Only studies published in English and full texts will be included in this review, with all databases searched from their inception to the current date (2013).

The databases to be searched will be JBI COnNECT+, Cochrane, Pubmed, CINAHL, SPORTSDiscus, OT Seeker, PEDro, and physiotherapy journals in the UK and Australia.

The search for unpublished studies or gray literature will include Mednar.

Initial key words to be used will be ‘dynamometry’, ‘dynamometer’, ‘manual muscle test’, ‘spinal cord injury’ and ‘muscle strength’.


Assessment of methodological quality

The quality evaluation tool Consensus-Based Standards for the Selection of Health Status Measurement Instruments (COSMIN) will be used to evaluate the methodological quality of the studies to be included in this review. The COSMIN tool has international acceptance and has been validated via research and therefore will provide a more robust approach for assessing clinimetric studies rather than the JBI tools.

The COSMIN tool has a four-point scored checklist (excellent, good, fair, poor) which consists of 10 separate sections, each dealing with a different measurement property. This review will use the domains of reliability, validity, responsiveness and interpretability. The COSMIN checklist is a modular tool which means it is not necessary to complete the whole checklist (see Appendix I). Two reviewers will be used during this review process and both will be familiar with the COSMIN tool and checklist (education via the 56-page COSMIN handbook). Any disagreements between the two reviewers will be resolved through discussion or via a third reviewer.
Data collection
Quantitative data will be extracted from studies using the attached data extraction form which has been developed specifically for this review (Appendix II)

Data synthesis
If appropriate a meta-analysis will be conducted.
If a meta-analysis is not appropriate, the narrative text will include raw data as presented in the included studies as well as the contextual data. Tables and figures will also be used in the presentation of results.

Conflicts of interest
It is declared that there are no conflicts of interest.

Acknowledgements
Our appreciation goes to the following Joanna Briggs Institute staff for their assistance and support:
Dr Suzanne Robertson-Malt, BHSc (Hons), PhD
Dr Catalin Tufanaru, MD, MPH
Dr Alla Grynevych, MD,MMSc
References


14. Dobson, F, Choi, YM, Hall, M, Hinman, RS. Clinimetric properties of observer-assessed impairment tests used to evaluate hip and groin impairments: A systematic review, Arthritis care and research. 2012; vol.64 (10); 1565-1575.

Appendix I: Appraisal instruments

The COSMIN checklist

Contact
CB Terwee, PhD
VU University Medical Center
Department of Epidemiology and Biostatistics
EMGO Institute for Health and Care Research
1081 BT Amsterdam
The Netherlands
Websites: www.cosmin.nl, www.emgo.nl
E-mail: cb.terwee@vumc.nl

Step 1. Evaluated measurement properties in the article

(The COSMIN checklist is a modular tool which means that it is not necessary to complete the whole checklist when evaluating the quality of a particular study. The following includes the COSMIN boxes to be used.)

Reliability
Box B
Measurement error
Box C
Content validity
Box D
Criterion validity
Box H
Responsiveness
Box I
Interpretability
Box J

Step 2. Determining if a study meets the standards for good methodological quality
Box B. Reliability: relative measures (including test-retest reliability, inter-rater reliability and intra-rater reliability)

Design requirements

1. Was the percentage of missing items given?  
   yes ☐  no ☐

2. Was there a description of how missing items were handled?  
   y ☐  n ☐

3. Was the sample size included in the analysis adequate?  
   y ☐  n ☐  ? ☐

4. Were at least two measurements available?  
   y ☐  n ☐

5. Were the administrations independent?  
   y ☐  n ☐  ? ☐

6. Was the time interval stated?  
   y ☐  n ☐

7. Were patients stable in the interim period on the construct to be measured?  
   y ☐  n ☐  ? ☐

8. Was the time interval appropriate?  
   y ☐  n ☐  ? ☐

9. Were the test conditions similar for both measurements? e.g. type of administration, environment, instructions  
   y ☐  n ☐  ? ☐

10. Were there any important flaws in the design or methods of the study?  
    y ☐  n ☐

Statistical methods

11. For continuous scores: Was an intraclass correlation coefficient (ICC) calculated?  
    y ☐  n ☐  N/A ☐

12. For dichotomous/nominal/ordinal scores: Was kappa calculated?  
    y ☐  n ☐  N/A ☐

13. For ordinal scores: Was a weighted kappa calculated?  
    y ☐  n ☐  N/A ☐  ? ☐

14. For ordinal scores: Was the weighting scheme described? e.g. linear, quadratic  
    y ☐  n ☐  N/A ☐
Box C. Measurement error: absolute measures

Design requirements

1 Was the percentage of missing items given?
   y ☐ n ☐

2 Was there a description of how missing items were handled?
   y ☐ n ☐

3 Was the sample size included in the analysis adequate?
   y ☐ n ☐ ? ☐

4 Were at least two measurements available?
   y ☐ n ☐

5 Were the administrations independent?
   y ☐ n ☐ ? ☐

6 Was the time interval stated?
   y ☐ n ☐

7 Were patients stable in the interim period on the construct to be measured?
   y ☐ n ☐ ? ☐

8 Was the time interval appropriate?
   y ☐ n ☐ ? ☐

9 Were the test conditions similar for both measurements? e.g. type of administration, environment, instructions
   y ☐ n ☐ ? ☐

10 Were there any important flaws in the design or methods of the study?
   y ☐ n ☐

Statistical methods

11 for CTT: Was the Standard Error of Measurement (SEM), Smallest Detectable Change (SDC) or Limits of Agreement (LoA) calculated?
   y ☐ n ☐

Box D. Content validity (including face validity)

General requirements

1 Was there an assessment of whether all items refer to relevant aspects of the construct to be measured?
   y ☐ n ☐ ? ☐

2 Was there an assessment of whether all items are relevant for the study population? (e.g. age, gender, disease characteristics, country, setting)
   y ☐ n ☐ ? ☐

3 Was there an assessment of whether all items are relevant for the purpose of the measurement instrument? (discriminative, evaluative, and/or predictive)
   y ☐ n ☐ ? ☐

4 Was there an assessment of whether all items together comprehensively reflect the construct to be measured?
   y ☐ n ☐ ? ☐
5 Were there any important flaws in the design or methods of the study?
☐ y ☐ n

**Box H. Criterion validity**

**Design requirements**

1 Was the percentage of missing items given?
☐ y ☐ n

2 Was there a description of how missing items were handled?
☐ y ☐ n

3 Was the sample size included in the analysis adequate?
☐ y ☐ n ☐ ?

4 Can the criterion used or employed be considered as a reasonable ‘gold standard’?
☐ y ☐ n ☐ ?

5 Were there any important flaws in the design or methods of the study?
☐ y ☐ n

**Statistical methods**

6 for continuous scores: Were correlations, or the area under the receiver operating curve calculated?
☐ y ☐ n ☐ N/A ☐

7 for dichotomous scores: Were sensitivity and specificity determined?
☐ y ☐ n ☐ N/A ☐

**Box I. Responsiveness**

**Design requirements**

1 Was the percentage of missing items given?
☐ y ☐ n

2 Was there a description of how missing items were handled?
☐ y ☐ n

3 Was the sample size included in the analysis adequate?
☐ y ☐ n ☐ ?

4 Was a longitudinal design with at least two measurement used?
☐ y ☐ n

5 Was the time interval stated?
☐ y ☐ n

6 If anything occurred in the interim period (e.g. intervention, other relevant events), was it adequately described?
☐ y ☐ n

7 Was a proportion of the patients changed (i.e. improvement or deterioration)?
☐ y ☐ n

---

doi: 10.11124/jbisrir-2013-969
Design requirements for hypotheses testing

For constructs for which a gold standard was not available:

8 Were hypotheses about changes in scores formulated a priori (i.e. before data collection)?
   y☐ n☐ ?☐

9 Was the expected direction of correlations or mean differences of the change scores of the instruments included in these hypotheses?
   y☐ n☐ N/A☐

10 Were the expected absolute or relative magnitude of correlations or mean differences of the change scores of the instruments included in these hypotheses?
   y☐ n☐ N/A☐

11 Was an adequate description provided of the comparator instrument(s)?
   y☐ n☐

12 Were the measurement properties of the comparator instrument(s) adequately described?
   y☐ n☐

13 Were there any important flaws in the design or methods of the study?
   y☐ n☐

Statistical methods

14 Were design and statistical methods adequate for the hypotheses to be tested?
   y☐ n☐ ?☐

Design requirement for comparison to a gold standard

For constructs for which a gold standard was available:

15 Can the criterion for change be considered as a reasonable gold standard?
   y☐ n☐ ?☐

16 Were there any important flaws in the design or methods of the study?
   y☐ n☐

Statistical methods

17 for continuous scores: Were correlations between change scores, or the area under the Receiver Operator Curve (ROC) curve calculated?
   y☐ n☐ N/A☐

18 for dichotomous scales: Were sensitivity and specificity (changed versus not changed) determined?
   y☐ n☐ N/A☐
Box J. Interpretability

1 Was the percentage of missing items given?
y☐ n☐

2 Was there a description of how missing items were handled?
y☐ n☐

3 Was the sample size included in the analysis adequate?
y☐ n☐ ?☐

4 Was the distribution of the (total) scores in the study sample described?
y☐ n☐

5 Was the percentage of the respondents who had the lowest possible (total) score described?
y☐ n☐

6 Was the percentage of the respondents who had the highest possible (total) score described?
y☐ n☐

7 Were scores and change scores (i.e. means and SD) presented for relevant (sub) groups? e.g. for normative groups, subgroups of patients, or the general population
y☐ n☐

8 Was the minimal important change (MIC) or the minimal important difference (MID) determined?
y☐ n☐

9 Were there any important flaws in the design or methods of the study?
y☐ n☐

Step 3: Determining the Generalisability of the results

Box Generalisability

Was the sample in which the instrument was evaluated adequately described? In terms of:

1 median or mean age (with standard deviation or range)?
y☐ n☐

2 distribution of sex?
y☐ n☐

3 important disease characteristics (e.g. severity, status, duration) and description of treatment?
y☐ n☐ N/A☐

4 setting(s) in which the study was conducted? e.g. general population, primary care or hospital/rehabilitation care
y☐ n☐

5 countries in which the study was conducted?
6 Language in which the instrument was evaluated?

7 Was the method used to select patients adequately described? e.g. convenience, consecutive, or random

8 Was the percentage of missing responses (response rate) acceptable?

Appendix II: Data extraction instruments

The following data extraction tool has been specifically designed using a combination of the data extraction worksheets used in a critical review by Stark et al.\textsuperscript{8} and clinimetric studies by Dobson et al.\textsuperscript{14} and de Koning et al.\textsuperscript{15}

<table>
<thead>
<tr>
<th></th>
<th>Study one</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Authors, year and ref.</strong></td>
<td></td>
</tr>
<tr>
<td><strong>Types of participants including diagnosis specifics</strong></td>
<td>Age</td>
</tr>
<tr>
<td></td>
<td>Gender</td>
</tr>
<tr>
<td></td>
<td>No. of participants</td>
</tr>
<tr>
<td></td>
<td>Diagnosis and co-morbidities</td>
</tr>
<tr>
<td></td>
<td>ASIA score (if known)</td>
</tr>
<tr>
<td></td>
<td>Time since injury</td>
</tr>
<tr>
<td><strong>Method of muscle strength test used including name of instrument.</strong></td>
<td>Manual muscle test</td>
</tr>
<tr>
<td></td>
<td>Isokinetic Dynamometry</td>
</tr>
<tr>
<td></td>
<td>Hand Held Dynamometry</td>
</tr>
<tr>
<td><strong>Trained examiner (yrs of experience)</strong></td>
<td></td>
</tr>
<tr>
<td><strong>Blinded examiner?</strong></td>
<td></td>
</tr>
<tr>
<td><strong>Clinimetric domain evaluated</strong></td>
<td>Reliability</td>
</tr>
<tr>
<td></td>
<td>Validity</td>
</tr>
<tr>
<td></td>
<td>Responsiveness</td>
</tr>
<tr>
<td></td>
<td>Interpretability</td>
</tr>
<tr>
<td><strong>Statistical analysis used</strong></td>
<td></td>
</tr>
<tr>
<td><strong>Results</strong></td>
<td></td>
</tr>
<tr>
<td><strong>Authors Conclusions</strong></td>
<td></td>
</tr>
<tr>
<td><strong>Reviewers conclusions including COSMIN score</strong></td>
<td></td>
</tr>
</tbody>
</table>