

Clinical assessment of hand oedema: a systematic review.

Short title: Assessment of hand oedema

Authors: Leanne K Miller¹, Christina Jerosch-Herold¹, Lee Shepstone²

¹ School of Health Sciences, University of East Anglia, Norwich

² Norwich Medical School, University of East Anglia, Norwich, UK

Corresponding author:

Leanne K Miller

The Queen's Building

School of Health Sciences

University of East Anglia

Norwich

NR4 7TJ

Leanne.miller@uea.ac.uk

Tel 01603 597206

Conflict of interest: The Authors declares that there is no conflict of interest.

Funding: Leanne Miller is funded by a National Institute for Health Research and Health Education England Clinical Doctoral Research Fellowship (CDRF-2014-05-064). Christina Jerosch-Herold is funded by a National Institute for Health Research Senior Research Fellowship (SRF-2012-05-119). This article presents independent research funded by the National Institute for Health Research (NIHR) and Health Education England. The views expressed are those of the authors and not necessarily those of the NHS, the NIHR or the Department of Health.

Informed consent: Not applicable.

Guarantor: LM

Ethical approval: Not applicable.

Contributorship: LM researched literature and conceived the review. LM, CJH and LS were involved in protocol development, data analysis and assisting with

manuscript drafting. All authors reviewed and edited the manuscript and approved the final version of the manuscript

Clinical assessment of hand oedema: a systematic review.

Abstract

Introduction

Assessment of oedema after trauma or surgery is important to determine whether treatment is effective and to detect change over time. Volumetry is referred to as the “gold standard” method of measuring volume. However this has practical limitations and other methods are available. The aim of this systematic review was to evaluate the psychometric properties of alternative methods used to assess hand oedema.

Methods

A search of electronic bibliographic databases was undertaken for any studies published in English reporting the psychometric evaluation of a method for measuring hand oedema, in an adult population with hand swelling from surgery, trauma or stroke. The Consensus-based Standards for the Selection of health Measurement Instruments checklist (COSMIN) was used to evaluate the methodological quality.

Results

Six studies met the inclusion criteria. Three methods were identified assessing hand oedema: perometry, visual inspection and the figure of eight tape measure, all were compared to volumetry. Four different psychometric properties were assessed. Studies scored fair or poor on COSMIN criteria. There is low quality evidence supporting the use of the figure of eight tape measure to assess hand volume. The perometer systematically overestimated volume and visual estimation had poor sensitivity and specificity.

Discussion

The figure of eight tape measure is the best alternative to volumetry for hand oedema. Benefits include reduced cost and time whilst having comparable reliability to the “gold standard”. Further research is needed to compare methods in patients with greater variability of conditions and with isolated digit oedema. Visual estimation of hand oedema is not recommended.

Keywords: Hand, oedema, assessment, outcome measures, volume

Introduction

Prolonged swelling has an impact on joint range of motion, soft tissue mobility, quality of scar tissue formation, function, strength, and aesthetics of the hand. These factors may delay a patient's recovery, return to work and usual activities of daily living and require frequent or increased out-patient appointments. ^[1]

Assessment of hand oedema after stroke, surgery or trauma offers valuable information to the treating therapist about the effectiveness of oedema management interventions, adherence to home therapy programmes ^[2] and activity levels. Objective measures are particularly important in the current economic climate to ensure that interventions and therapy time can be justified. For this reason measures need to not only be reliable but also responsive to detect clinically important change over time. Whilst it is best practice to maintain consistency of therapists between treatment sessions, in busy clinics and regional units patients are often seen by multiple therapists across their episode of care and therefore assessment tools are needed with a high level of inter- and intra-rater reliability.

The volumeter, which uses Archimedes' principle of water displacement ^[3], has been in existence since the 1950's ^[4] however its usage in therapy departments appears to be reducing. This method has documented reliability and validity ^[2] and has a margin of error of less than 1% ^[5]. It is referred to as the 'gold standard' way of assessing hand volume when oedema is generalised to the hand and not isolated to a digit ^[6] however it is not always a feasible method, for example where immersion of the hand in water is contraindicated due to wounds or dressings. The volumeter kit is also expensive at approximately £300 and requires a lengthy set up to ensure the water in the volumeter is completely level and a constant water temperature is maintained ^[7-10]. Furthermore, consistency in positioning the hand and arm is essential and the need to maintain a still limb may also exclude some patients ^[11]. Potential increases in pain from the dependent limb position and length of time to allow all displaced water to be collected are further limitations ^[5]. The volumeter is often impractical in busy clinic settings where space is limited and frequent hand oedema assessments need to be performed or in patients who have focal swelling limited to a single digit.

Alternative methods include visual inspection of the oedematous hand and documenting a grade using terminology acceptable to that department such as mild, moderate and severe for example. This subjective assessment of hand volume is based on colour and tautness of the skin and appearance of defined anatomical landmarks or lack thereof. Due to varying perceptions of severity between clinicians and difficulties with recall between sessions with the same clinician, visual inspection alone may not be sufficient to give an accurate measurement of hand volume and an objective measurement of oedema needs to be performed.

Another alternative which is quicker and cheaper is using a tape measure in a circumferential or figure of eight method. This technique is simple and reproducible if

used with standardized landmarks and can be used in the presence of wounds. The limitation with the figure of eight method is its exclusion of the digits so this may not be the method of choice to use in cases of isolated digital swelling as the placement of the tape around the wrist and palm only measures the volume of the regions covered by the tape and does not include digits.

Other methods of determining volume exist such as 3D laser scanners [12-14], 3D camera [15] and perometer [16] (an infrared optoelectric measuring device). Whilst these methods are not routinely used by hand therapists to measure oedema, information on their application and psychometric properties could be transferable to use in clinical practice on the hand. The hand presents a unique challenge when measuring volume due to its shape and structure and this may mean some methods are not suitable to use.

In light of the information presented above the rationale for conducting this systematic review was to establish which oedema assessment method has the strongest psychometric evidence.

The objective of this systematic review were to:

1. Establish the current quantity and quality of evidence on tools designed to assess hand oedema
2. Evaluate the psychometric properties of these tools
3. Identify factors affecting the standardisation of these tools.

Methods

We conducted a systematic review using PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analysis) recommendations [17].

The following electronic bibliographic databases were searched: The Cochrane Library (Wiley InterScience), MEDLINE (via Ovid), EMBASE (via Ovid), AMED (via Ovid), CINAHL (via EBSCO), SPORTDiscus (via EBSCO), PEDro (Physiotherapy Evidence Database) - Allied Health Evidence. Trial registers (Cochrane Central Register of Controlled Trials [CENTRAL] and WHO International Clinical Trials Registry Platform) from inception to March 2017 were searched using the terms: 'Hand/', 'Edema/', 'Hand' adj 'size', 'hand' adj 'volume', 'perometer'. Additional studies were searched for by examining the reference list of retrieved studies.

Eligibility

Criteria for inclusion were: English language publications reporting psychometric evaluation of an assessment to measure hand volume in an adult population with hand oedema. Eligible forms of hand oedema were following surgery or trauma or from a disease or condition affecting the hand irrespective of any treatment given

(e.g. stroke, lymphoedema), where hand oedema measurements are expressed as volume (ml), girth or circumference (cm/mm) or as a severity description.

Studies were excluded if the psychometric evaluation had been completed on healthy participants only, animal studies, studies which assessed the upper limb and forearm in addition to the hand and studies where oedema was investigated at an organ or cellular level.

Screening

One reviewer (LM) read the titles of all citations retrieved from electronic database searches and removed all citations which were not related to the assessment of hand oedema. Abstracts of the remaining articles were screened to check for eligibility by one reviewer (LM). Full text articles were obtained for all abstracts meeting the inclusion criteria.

Data Extraction

Data extraction of included studies was done by the lead author (LM) using a purposely designed data extraction form. This form summarized details on study design, sample, interventions, outcomes, and results. On occasions when there was doubt over the interpretation of the data being extracted a second reviewer (CJH) also complete the data extraction independently using the same form to verify understanding and clarity of extracted data.

Assessment of methodological quality

The Consensus-based Standards for the selection of health Measurement Instruments checklist (COSMIN) ^[18] was used to evaluate the methodological quality of the studies. This checklist was originally designed for use in Health Related Patient Reported Outcomes (HR-PRO) but can be used to evaluate other kinds of health measurement instruments such as performance-based tests and clinical rating scales. The COSMIN checklist is made of nine domains relating to different psychometric properties. Each study was assessed using the relevant domain for the psychometric property being evaluated i.e. reliability, validity or responsiveness by the primary reviewer (LM). The second reviewer (CJH) completed the checklist for two of the six included studies and the agreement between the reviewers was checked to ensure consistent grading across each domain for each study. There was 86% agreement between primary and secondary reviewer on the selected two studies, the inconsistencies in scores were settled with discussion and resulted in 100% agreement. Each domain has between 7 and 14 questions which are graded on a four-point rating scale: 'excellent', 'good', 'fair' or 'poor' according to the descriptors given under each category. The lowest score counts method is recommended to give an overall quality judgement.

Included studies were grouped according to the assessment method used: figure of eight, perometry and visual inspection. This formed the basis of how results were

reported. Meta-analysis was not possible because of heterogeneity in assessment tools, methods or reporting of results.

Results

Six studies met the inclusion criteria (see Figure 1 for PRISMA flow diagram) and were included in this review.

A total of 243 participants were included in the 6 studies, with sample sizes ranging from 24 to 88. Participants had a range of musculoskeletal injuries, burns, lymphoedema, post orthopaedic surgery or cerebrovascular accident (CVA). Only one study ^[19] used a healthy comparison group when assessing the reliability of the perometer in women with and without lymphedema.

Three methods of assessing oedema were used: figure of eight tape measure, perometer and visual observations by clinicians. All were compared with volumetry as the “gold standard” method as this has excellent intra- and inter-rater reliability (ICC 0.99, respectively). ^[20]

Four studies ^[20, 21, 22, 23] assessed the reliability of the figure of eight comparing it to the volumeter, however not all statistical results were reported. Leard ^[23] also assessed the responsiveness of these two methods of assessing oedema.

One study ^[24] assessed the reliability of using visual inspection compared to volumetry, and the one study ^[19] evaluated the reliability of the perometer compared to the volumeter.

Four studies ^[19, 20, 21, 22] assessed criterion-related validity and, along with Leard ^[23], also investigated measurement error of their respective oedema assessment tools. See Table 1 for an overview of the studies and the psychometric properties they assessed.

The results are presented according to the measurement tool used. Tables 2 to 5 show the quality rating table for each psychometric property/ study using the COSMIN checklist.

[insert Figure 1 and Tables 1 to 5 near here]

Perometer:

Lee et al ^[19] assessed 20 women with and 20 women without lymphedema of the hand and reported reliability data both for subgroups and the whole group. Excellent inter- and intra-rater reliability was demonstrated for the perometer (ICC= 0.99; 95% CI 0.98-0.99 and ICC= 0.99; 95% CI 0.98-0.99, respectively). Similarly, excellent inter- and intra-rater reliability (ICC > 0.99) was observed for the two subgroups. There was no significant difference between measurements taken by different raters or between the two measurements taken by tester 1. Whilst Lee et al ^[19] gave confidence intervals with their ICC's they did not report the standard error of

measurement (SEM) which gives an absolute index of reliability rather than a relative measure of reliability.

However, the perometer systematically overestimated hand volume by a mean of 24ml compared with the volumeter. Mean hand volume (n=20 women without lymphoedema) is 380ml which equates to a 6% overestimation in volume. Whilst the perometer has excellent inter- and intra-rater reliability comparable to the gold standard volumeter and a very good concordance correlation, calibration issues led to a 6% overestimation and therefore the two methods for measuring hand volume should not be used interchangeably.

Lee ^[19] commented on the potential issue of the perometer being its inability to discriminate interdigital spaces and therefore it interprets this space as volume and includes it in the overall volume measurement. It may also be difficult for some patients to maintain a static position over the period required to complete the assessment and therefore a slight shift of the hand may also result in an overestimation of the actual volume.

This study ^[19] scored 'fair' overall across absolute error, reliability and criterion validity categories of the COSMIN quality assessment.

Visual Inspection:

Visual observations were carried out by experienced therapists during a 1 hour consultation for post-stroke arm/hand problems. The therapists classified the amount of hand swelling observed during visual inspection as being nil, minor or severe. Post et al ^[24] assessed 88 hands after their first stroke. Whilst the authors claim there was "a clear relationship between the assessment by the physical therapists and the adjusted volume scores" (mean volumeter scores were adjusted from the population data), the results actually indicate a lack of agreement between clinical and volumetric assessment of oedema. A 67% agreement was found between classification of oedema by therapists and the volumeter. A Kappa value of 0.34 highlights a fair level of agreement. However no confidence intervals were provided.

Although Post et al ^[24] did not report sensitivity and specificity, these have been calculated from the data provided. Calculations were completed by authors LM and CJH. Sensitivity of visual inspection by therapists was 74% indicating that in 26 patients therapists missed oedema using this technique. In 76% (22/29) of cases the therapist reported oedema, the volumeter also agreed. Therapists' clinical judgement classified only 4.5% (n=4) of the group as having major oedema when the volumeter results show that actually 18.5% of the group were in this category.

Specificity of visual inspection was 63%, meaning that in 63% (37/44) of cases the therapist reported no swelling, the volumeter also agreed. Therapists' clinical judgement classified 40% of the population (n=44) as having no oedema whereas the volumeter results indicate only 2.2% of the group had no oedema.

This study scored 'fair' on the COSMIN quality assessment in both criterion validity and reliability categories

Across the two categories scores of fair, good or excellent were given for each question. However, in light of the lack of sensitivity and specificity calculations this brought the overall rating down to poor.

Figure of Eight Tape Measure:

There were slight variations in the methods used to administer the figure of eight assessment between the 4 studies [20, 21, 22, 23] and often some details were not adequately documented.

Leard's [23] paper reports completing intra-rater reliability assessment for the figure of eight, however actually only documents inter-rater reliability results.

Intraclass Correlation Coefficients (ICC) for intra-rater reliability ranged between 0.89 and 0.99 across the 3 studies (Leard [23] did not report intra-rater reliability) demonstrating excellent levels of intra-rater reliability with the figure of eight method. Standard Error of the Mean (SEM) ranged between 0.28-0.70cm across the three studies [20, 22, 23] which documented this.

High inter-rater reliability was also demonstrated across the 4 studies with an ICC range of 0.84 - 0.99, and SEM range of 0.28-0.71 cm. The study which reported the highest ICC of 0.99 [20] also reported the smallest SEM of 0.28 cm, and the same was true for the reverse of this, 0.86 ICC and 0.71 cm SEM [22, 23].

Leard [23] also assessed the responsiveness of the figure of eight compared to the volumeter which demonstrated similarly small effect sizes (ES) (ES=0.26 for figure of eight and ES=0.19 for volumeter) highlighting that the ability of the tools to detect changes in hand volume over time is comparable but slightly favours the figure of eight. When reporting the standardized response mean (SRM) however, the figure of eight had a slightly lower value (SRM=0.87) than the volumeter (SRM=1.04) which contrasts with the effect sizes. As no summary statistics were given, we are unable to replicate the analysis to verify these results.

Of the four studies which used the figure of eight, two scored poor [22, 23] and two fair [20, 21] in the COSMIN quality evaluation tool.

Discussion

The aims of this systematic review were to review the quality and quantity of current evidence on the psychometric properties of methods for assessing hand oedema and identify factors which may affect the standardisation of these methods when used on the hand. A discussion of the findings and implications for practice will be presented in this section.

The review found limited low quality evidence to support the use of the figure-of-eight tape measure to assess hand volume in patients with acute or chronic oedema from a traumatic, lymphatic or neurological cause.

Whilst the perometer had similar levels of reliability to that of the “gold standard” volumeter it showed a systematic overestimation which equated to 6% of total hand volume highlighting its incompatibility to be used interchangeably with the volumeter. Issues around hand position and accuracy of the infrared beam to discriminate hand volume and space contributed to the overestimation of hand volume.

Visual inspection had a fair level of agreement with the volumeter. However results show that visual inspection may miss some patients with oedema and wrongly diagnose some patients as having oedema.

Assessment of methodological quality

The COSMIN ^[18] checklist was used to assess the methodological quality of the studies. It was developed specifically to assess health-related patient rated outcome measures (HR-PRO) These scales or questionnaires are often made up of several items designed to measure a latent construct. Therefore some sections and questions of the checklist are not appropriate when evaluating measures of a single domain such as hand volume.

The current scoring system works on a 4-point rating scale: excellent, good, fair and poor. This was adapted from a dichotomous response option (yes/no) and accounts for some of the issues with scoring. In the majority of questions there are descriptors under each rating which qualifies what the paper must report in order to achieve that rating. However, in some cases, descriptors have not been included.

In these cases the missing ‘good’ and ‘fair’ descriptions were appropriate as the question related to the completion of statistical tests which warrant only a yes (excellent) or no (poor) answer. However in some instances the gap or difference between descriptors seemed arbitrary and often it is difficult to find the most appropriate score based on the descriptions given to accurately reflect the quality of the paper. The working group who developed the 4-point rating scale report, that for some questions, it was not possible to define four different response options

A worst score counts method is used to give an overall quality rating for each measurement property. A poor score on any one item is thus considered to represent a fatal flaw ^[25]. Other methods of scoring have been considered ^[25, 26] and whilst the overall score is often lower than the subjective judgement of the marker, this method has been agreed, following a Delphi consensus study ^[26] to be the most appropriate. The scoring method however is arbitrary and the validity and reliability of the current recommended scoring system has not been investigated ^[25]. Despite the limitations of this critical evaluation tool, it is the only standardized rating tool which can be applied to health-related clinician-derived measurement instruments.

Sample Size

Four studies [19, 20, 21, 24] scored 'fair' in all measurement properties assessed. Borthwick [22] and Leard [23] scored poor across all three measurement properties assessed (reliability, criterion validity and measurement error). Both studies scored 'poor' based on a single item - adequate sample size. Indicative sample sizes are given as a guide for each response option based on a "rule of thumb" [25] however authors report that definitions of an 'adequate' sample size may differ depending on the situation and that markers should have the flexibility to adapt the scoring system based on their own application. This explains why certain items do not have specific criteria, such as the time between assessments in test-retest evaluation. Whilst this flexibility is useful to ensure the scoring system is representative of a particular instrument and its setting, it may cause issues regarding the standardisation of the checklist's scoring system and comparison between markers and across papers.

Factors affecting standardisation

Perometer

Incorrect limb position has been described as the main reason for the poor accuracy of the volume measurement obtained by the perometer. This has been previously documented [27-29]. Stanton [27] reports that large measurement errors occurred when the limb was not perpendicular to the laser beam. Lee [19] attempted to reduce measurement error arising from limb position by ensuring all patients held their digits tightly together including the thumb close against the index finger. The perometer however viewed the hand as an elliptical object and included interdigital air spaces as tissue and therefore this was included in the overall volume.

Inter- and intra-rater reliability was lower for the sub-group of 20 women without lymphoedema in this study. When a hand is swollen (such as in lymphoedema) it takes on more of an triaxial ellipsoid shape and thus the laser beams cannot detect the diminished or absent interdigital air spaces resulting in greater reliability measures for patients with swelling than those without.

Lee [19] highlights that the perometer has advantages over the water displacement method in that it can be used on patients with skin conditions and open wounds where using the volumeter may not be feasible. It is much quicker to administer and requires less set up time however, the measurement errors described above are not isolated to the hand. Man [30] reports that the angle of the knee could affect the volume measure by up to 11% using the perometer. It is possible that even with a standardized protocol and limb position, the unique position of the thumb in a frontal plane makes optoelectric imaging unsuitable for use on the hand when assessing volume. Whilst a lightweight and portable version of the perometer exists, the standard version would require a permanent space in a clinically setting and costs between £10,000-15,000 depending on the model.

Figure of eight

The type of tape measure may also affect the accuracy of the measurements obtained. Retractable measures may have more 'give' to them and can be pulled tighter. Particularly in oedematous hands the danger is that whilst concentrating on locating anatomical landmarks to achieve accurate tape placement the tension being applied can actually displace oedematous tissue. Education, practice and standardised protocols for administration may reduce this risk, such as those provided by the American Society of Hand Therapists [31].

Timing of assessments

Post et al [24] highlight a limitation of their study as being the time between assessments. Median time between clinical evaluation and volumetric assessment was 7 days. They report that time between assessments did not influence results. However it was shown that visual inspection may underestimate the number of patients with oedema and overestimate the number of patients without oedema. As the clinical evaluation was performed first, the oedema could have improved spontaneously or worsened by the time the volumetric assessment took place 7 days later. The authors do not report what, if any, therapy interventions took place during the 7 days which may account for a change in volume. A higher level of agreement with clinical evaluation could have been observed if the volumetric assessments were completed at a more appropriate time, that is on the same day to the clinical evaluation.

Patient Rated Outcome Measures

To the best of the authors' knowledge there are no patient-rated outcome measures currently being used which assess or grade swelling from the patient's perception. Although oedema is an observable condition which can be measured by the clinician using a tape measure or volumeter, it is also a subjective condition, like pain, where a patient may feel pressure or tightness which limits full movement from oedema even if this swelling is not detectable to the eye. It would be useful to assess the relationship between a clinician-derived measure such as the figure of eight method or volumeter and a patient-rated outcome measure which grades their perception of the swelling. This could be a valuable and time efficient method of evaluating treatment effectiveness from the patient's perspective which could compliment clinician-derived assessments and help to establish a minimally important difference for specific diagnostic subgroups.

Location of oedema

Circumferential measurements may be the only option for measuring digital swelling however, in areas where bony landmarks don't exist such as the mid forearm, placement of the tape measure can vary between therapists even when the location has been documented. In the hand, Maihafer ^[5] argued that the figure of eight method is better able to capture hand volume than single joint or single plane measures, which don't adequately reflect volume or size, however their study used a healthy cohort with no hand oedema. Studies which have compared circumferential measures with the volumeter in lymphoedema patients with upper limb oedema have not included circumferential measurements of the hand ^[16, 32,33]. Previous studies investigating the psychometric properties of the figure of eight tape measure in comparison to the volumeter included patients with diverse hand and wrist trauma but often do not specify the exact location of oedema ^[20]. Whilst previous studies have reported the figure of eight tape measurement method is as reliable as the volumeter ^[6] these only used a healthy cohort without hand oedema and therefore the unique challenges of assessing a hand with increased fluid may not be captured.

Limitations of the review

This systematic review has a number of limitations. Firstly, the included studies focus on hand oedema only. Whilst methods such as volumetry, perometry and visual inspection will take into account swelling of the digits as well as the hand, the figure-of-eight method neglects the digits and therefore could not be used in isolated finger swelling. Circumferential measurements of digits which are used when assessing isolated digit swelling was not a method described in the selected papers.

The volumeter also includes volume of the wrist and distal forearm along with the hand and digits, whereas the figure-of-eight starts at the ulnar and radial styloid and does not take into account the presence of any swelling at the proximal wrist and distal forearm. The inclusion criteria for this systematic review specified hand oedema only, however, as the volumeter was used as the comparator in all studies, it is feasible, particularly in patients with lymphedema ^[19, 22], stroke ^[24] and burns ^[21] that the swelling extended into the arm and that this may have been included in volumetric assessment but not in the figure of eight measurements. It is also unclear from the literature where the exact cut-off point for the perometer's laser beam is on the hand or wrist and if the clinicians based their visual evaluation on the hand only or included the wrist or forearm.

Another limitation could be the generalisability of the results. Whilst it appears the results are generalisable to therapists with varying levels of experience, due to the limited number of papers meeting the inclusion criteria, the results may not be generalisable to patients with different hand conditions or in different settings such as chronic, rehabilitation or very acute phase of oedema.

Conclusion

Based on a review of the current evidence the figure of eight oedema assessment is the best alternative to the volumeter. It has comparable reliability to the current gold standard, the volumeter. However replicating studies with a larger number of participants with greater variability of conditions is needed. The perometer is expensive and prone to measurement errors resulting in exaggerated oedema measurements. Many departments may not have access to a volumeter and the submersion of the hand may not be a feasible option in the presence of wounds or dressings. However, the temporary removal or reduction of dressings to assess oedema with a tape measure is a feasible option which offers therapists a quick, cheap, and simple method of objectively assessing hand volume. The use of a protocol is recommended to increase inter- and intra-rater reliability. Visual estimations should be avoided given the poor intra- and inter-rater reliability and correlation with objective measures.

References

1. Miller L, Jerosch-Herold C, Shepstone L. Effectiveness of edema management techniques for subacute hand edema: A systematic review. *Journal of Hand Therapy*. 2017. In press
2. Farrell K, Johnson A, Duncan H et al. The intertester and intratester reliability of hand volumetrics. *Journal of Hand Therapy*. 2003; 16:292-9
3. Finnerty G, Corbet T. *Hydrotherapy*. New York: Frederick Ungar Publishing Co, 1960
4. Eccles MV, Hand Volumetrics. *BR J Phys Med*. 1956; 19, 5-8
5. Maihafer G, Llewellyn M, Pillar W et al. A comparison of the Figure-of-eight Method and Water Volumetry in Measurement of Hand and Wrist Size. *Journal of Hand Therapy*. 2003;16:305-310
6. Pellachia G. Figure-of-eight Method of Measuring Hand Size. Reliability and Concurrent Validity. *Journal of Hand Therapy*. 2003; 16 300-404
7. King T. The Effect of Water Temperature on Hand Volume during Volumetric Measurement Using the Water Displacement Method. *Journal of Hand Therapy*. 1993, 202-204.
8. Waylett-Rendall J, Siebly DS. A study of the accuracy of a commercially available volumeter. 1991. *J Hand Therapy* 4:10-13

9. Brand PW. *Clinical Mechanics of the Hand*. 1985. St. Louis. C V Mosby.
10. Stern EB. Volumetric comparison of seated and standing test postures. *Am J Occup Ther*. 1991; 45: 801-4.
11. DeVore GL, Hamilton GF. Volume measuring of the severely injured hand. *Am J Occup Ther* 1968. 22:16-18.
12. Harrison J, Nixon M, Fright W et al. Use of hand-held laser scanning in the assessment of facial swelling: a preliminary study. *British Journal of Oral and Maxillofacial Surgery*. 2004;42, 8-17.
13. Kau CH, Cronin A, Durning P et al. A new method for the 3D measurement of postoperative swelling following orthognathic surgery. *Orthod Craniofacial Res*, 2006; 9: 31-37.
14. Mestree S, Veye F, Perez-martin A et al. Validation of lower limb segmental volumetry with hand=hand, self-positioning three-dimensional laser scanner against water displacement. *Journal of Vascular Surgery Venous and Lymphatic Disorders*. 2014; 2:39-45
15. Yip E, Smith A, Yoshino M et al. Volumetric evaluation of facial swelling utilizing a 3-D range camera. *International Journal of Oral and Maxillofacial Surgery*. 2004; 33: 179-182.
16. Deltombe T, Jamart J, Recloux S, Legrand C, Vandembroeck N, Theys S, Hanson P. Reliability and limits of agreement between water displacement and optoelectric volumetry in the measurement of upper limb oedema. *Lymphology* 2007, 40, 26-34
17. David Moher, PhD; Alessandro Liberati, MD, DrPH; Jennifer Tetzlaff, BSc; Douglas G. Altman, DSc, and the PRISMA Group. Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. *Ann Intern Med*. 2009;151(4):264-269
18. Mokkink L, Terwee C et al. COSMIN checklist manual. January 2012. Accessed on <http://www.cosmin.nl/images/upload/files/COSMIN%20checklist%20manual%20v9.pdf>

19. Lee M-J, Boland R, Czerniec S, Kilbreath S. Reliability and Concurrent Validity of the Perometer for Measuring Hand Volume in Women With and Without Lymphedema. *Lymphatic research and Biology*. 2011 9,1;13-18
20. Leard J, Breglio L, Fraga L et al. Reliability and Concurrent Validity of the Figure of Eight Method of Measuring Hand Size in Patients with Hand Pathology. *Journal of Orthopaedic and Sports Physical Therapy*. 2004.34(6); 335-240.
21. Dewey W, Hedman T, Chapman T et al. The Reliability and Concurrent Validity of the Figure-of-Eight Method of Measuring Hand Edema in Patients with Burns. *Journal of Burn Care Research*. 2007; 28:157-162
22. Borthwick Y, Paul L, Sneddon M et al. Reliability and validity of the figure-of-eight method of measuring hand size in patients with breast cancer-related lymphedema. *European Journal of cancer Care*. 2013 22; 196-201.
23. Leard J, Crane B, Mayette D et al. Responsiveness of the figure-of-eight tape measurement to detect hand size changes in patients with acute and chronic hand pathologies. *Hand Therapy*. 2008 13:3, 84-90
24. Post M, Visser-Meily J, Boomkamp-Koppen H et al. Assessment of oedema in stroke patients: comparison of visual inspection by therapists and volumetric assessment. *Disability and Rehabilitation*. 2003; 25,22: 1265-1270
25. Terwee CC, Mokkink LB, Knol D et al. Rating the methodological quality in systematic reviews of studies on measurement properties: a scoring system for the COSMIN checklist. *Qual Life Res*. 2012; 21:651–657
26. Mokkink, L. B., Terwee, C. B., Patrick, D. L. et al. The COSMIN checklist for assessing the methodological quality of studies on measurement properties of health status measurement instruments: An international Delphi study. *Quality of Life Research* 2010, 19, 539–549
27. Stanton AW, Northfield JW, Holyroyd B et al. Validation of an optoelectronic limb volumeter (Perometer). *Lymphology* 1997; 30: 77-97.
28. Hebeda CL, De Boer EM, Verburgh CA et al. Lower limb volume measurements: Standardization and reproducibility of an adapted Phlebology. 1993; 8: 162-166.

29. Louisy F, Schroiff P. Plethysmography with optoelectronic sensors: Comparison with mercury strain gauge plethysmography. *Aviat Space Environ Med* 1995; 66: 1191-1197
30. Man IOW, Elsabagh SM, Morrissey MC. The effect of different knee angles on knee volume measured with the perometer device in uninjured subjects, *Clin Physilo Funct Imaging* 2003; 23: 114.
31. Lavelle K, Breger Stanton D. American Society of Hand Therapists TM Key Recommendations for Outcome Evaluation of Edema Measurement of Edema in the Hand Clinic 2. Conceptual Basis for Testing Tests/Methods Used to Measure Edema. Available at <https://www.researchgate.net/publication/257927122>. Accessed on 19.6.17
32. Chen Y-W, Tsai HJ, Hung H-C, Tsauo J-Y. Reliability study of measurements for lymphedema in breast cancer patients. *American Journal of Physical Medicine and Rehabilitation* 2008, 87, 33-38
33. Gjorup C, Serahn B, Hendel H.W. Assessment of volume measurement of breast cancer-related lymphedema by three methods: circumference measurement, water displacement and dual action X-ray absorptiometry. *Lymphatic Research and Biology* 2010, 8, 111-119.

Table 1: Overview of included studies, cohort, assessment tool and psychometric properties assessed

Author	Pt type	Tools assessed	Psychometric properties assessed
Post 2003	88 hands post first CVA	Visual inspection Vs Volumeter	Reliability
Leard 2004	33 hands post trauma/surgery	Figure of Eight Vs Volumeter	Reliability, criterion validity, measurement error.
Dewey 2007	33 burned hands	Figure of Eight Vs Volumeter	Reliability, criterion validity, measurement error.
Leard 2008	25 hands post trauma/surgery	Figure of Eight Vs Volumeter	Reliability, responsiveness, measurement error.
Lee 2011	20 hands with and 20 hands without lymphedema	Perometer Vs Volumeter	Reliability, criterion validity, measurement error.
Borthwick 2011	24 hands with lymphedema	Figure of Eight Vs Volumeter	Reliability, criterion validity, measurement error.

Table 2: COSMIN Quality Assessment Table - Absolute error: absolute measures.

Study/ Question No.	1	2	3	4	5	6	7	8	9	10	11	“Worst score counts”
Leard 2004	N/A	N/A	Fair	Excellent	Excellent	Excellent	Good	Excellent	Excellent	Fair	Excellent	Fair
Leard 2008	N/A	N/A	Poor	Excellent	Excellent	Excellent	Good	Excellent	Excellent	Fair	Excellent	Poor
Dewey 2007	N/A	N/A	Fair	Excellent	Excellent	Excellent	Good	Excellent	Excellent	Fair	Excellent	Fair
Lee 2011	N/A	N/A	Fair	Excellent	Excellent	Excellent	Good	Excellent	Excellent	Fair	Excellent	Fair
Borthwick 2013	N/A	N/A	Poor	Excellent	Excellent	Excellent	Good	Excellent	Excellent	Fair	Excellent	Poor

- 1). Was the percentage of missing items given?
- 2). Was there a description of how missing items were handled?
- 3). Was the sample size included in the analysis adequate?
- 4). Were there at least 2 measurements?
- 5). Were the administrations independent?
- 6). Was the time interval stated?
- 7). Were the patients stable in the interim period on the construct to be measured?
- 8). Was the time interval appropriate?
- 9). Were the test conditions similar for both measurements? E.g type of administration, environment , instructions
- 10). Were there any important flaws in the design or methods of the study?
- 11). For CTT: Was the Standard Error of Measurement (SEM), Smallest Detectable Change (SDC) or Limits of Agreement (LoA) Calculated?

Table 3: COSMIN Quality Assessment Table - Reliability

Study/Question No.	1	2	3	4	5	6	7	8	9	10	11	12	13	14	"Worst score counts"
Leard (2008)	N/A	N/A	Poor	Excellent	Excellent	Excellent	Fair	Fair	Excellent	Fair	Excellent	N/A	N/A	N/A	Poor
Dewey (2007)	N/A	N/A	Fair	Excellent	Excellent	Fair	Fair	Fair	Excellent	Fair	Excellent	N/A	N/A	N/A	Fair
Post (2003)	N/A	N/A	Good	Excellent	Excellent	Excellent	Good	Fair	Fair	Fair	N/A	Excellent	Excellent	Good	Fair
Borthwick (2013)	N/A	N/A	Poor	Excellent	Excellent	Excellent	Good	Excellent	Excellent	Fair	Excellent	N/A	N/A	N/A	Poor
Leard (2004)	N/A	N/A	fair	Excellent	Excellent	Excellent	Good	Excellent	Excellent	Fair	Excellent	N/A	N/A	N/A	Fair
Lee (2011)	N/A	N/A	Fair	Excellent	Good	Fair	Good	Fair	Good	Fair	Excellent	N/A	N/A	N/A	Fair

1. Was the percentage of missing items given?
2. Was there a description of how missing items were handled?
3. Was the sample size included in the analysis adequate?
4. Were at least two measurements available?
5. Were the administrations independent?

6. Was the time interval stated?
7. Were patients stable in the interim period on the construct to be measured?
8. Was the time interval appropriate?
9. Were the test conditions similar for both measurements?
10. Were there any important flaws in the design or methods of the study?
11. For continuous scores: Was an intraclass correlation coefficient (ICC) calculated?
12. For dichotomous/nominal/ordinal scores: Was kappa calculated?
13. For ordinal scores: Was a weighted kappa calculated?
14. For ordinal scores: Was the weighting scheme described? e.g. linear, quadratic

Table 4: COSMIN Quality Assessment Table - Criterion Validity

Study/ Question No.	1	2	3	4	5	6	7	“Worst score counts”
Dewey (2007)	N/A	N/A	Fair	Excellent	Fair	Excellent	N/A	Fair
Post (2003)	N/A	N/A	Good	Excellent	Excellent	Excellent	Poor	Fair
Borthwick (2013)	N/A	N/A	Poor	Excellent	Fair	Excellent	N/A	Poor
Leard (2004)	N/A	N/A	Fair	Excellent	Fair	Excellent	N/A	Fair
Lee (2011)	N/A	N/A	Fair	Excellent	Fair	Excellent	N/A	Fair

1. Was the percentage of missing items given?
2. Was there a description of how missing items were handled?
3. Was the sample size included in the analysis adequate?
4. Can the criterion used or employed be considered as a reasonable ‘gold standard’?
5. Were there any important flaws in the design or methods of the study?
6. For continuous scores: Were correlations, or the area under the receiver operating curve calculated?
7. For dichotomous scores: Were sensitivity and specificity determined?

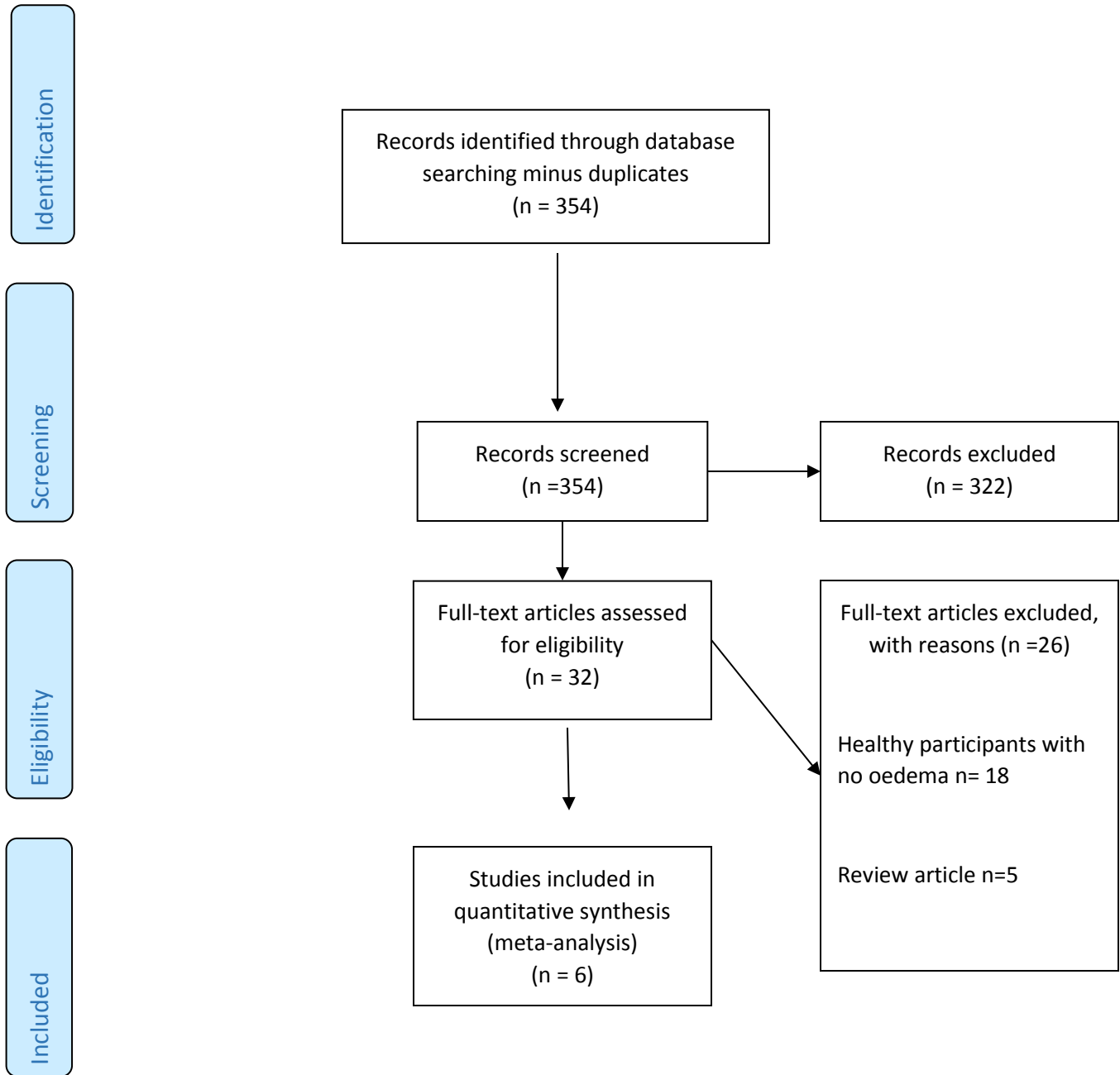
Table 5- COSMIN Quality Assessment Table – Responsiveness

Study/ Question No.	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	“Worst score counts”
Leard (2008)	N/A	N/A	Poor	Excellent	Excellent	Fair	Excellent	N/A	N/A	N/A	Excellent	Fair	Fair	Excellent	Excellent	Fair	Excellent	N/A	Poor

1. Was the percentage of missing items given?
2. Was there a description of how missing items were handled?
3. Was the sample size included in the analysis adequate?
4. Was a longitudinal design with at least two measurement used?
5. Was the time interval stated?
6. If anything occurred in the interim period (e.g. intervention, other relevant events), was it adequately described?
7. Was a proportion of the patients changed (i.e. improvement or deterioration)?
8. Were hypotheses about changes in scores formulated a priori (i.e. before data collection)?
9. Was the expected direction of correlations or mean differences of the change scores of HR-PRO instruments included in these hypotheses?
10. Were the expected absolute or relative magnitude of correlations or mean differences of the change scores of HR-PRO instruments included in these hypotheses?
11. Was an adequate description provided of the comparator instrument(s)?
12. Were the measurement properties of the comparator instrument(s) adequately described?
13. Were there any important flaws in the design or methods of the study?
14. Were design and statistical methods adequate for the hypotheses to be tested?
15. Can the criterion for change be considered as a reasonable gold standard?
16. Were there any important flaws in the design or methods of the study?
17. For continuous scores: Were correlations between change scores, or the area under the Receiver Operator Curve (ROC) curve calculated?
18. For dichotomous scales: Were sensitivity and specificity (changed versus not changed) determined?



Fig 1. PRISMA 2009 Flow Diagram



From: Moher D, Liberati A, Tetzlaff J, Altman DG, The PRISMA Group (2009). Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. PLoS Med 6(7): e1000097. doi:10.1371/journal.pmed1000097