# Measures of upper limb function for people with neck pain. A systematic review of measurement and practical properties

ALRENI, Ahmad Salah Eldin, HARROP, Deborah <http://orcid.org/0000-0002-6528-4310>, LOWE, Anna <http://orcid.org/0000-0001-5297-8957>, POTIA, Tanzila, KILNER, Karen <http://orcid.org/0000-0003-0196-8518> and MCLEAN, Sionnadh Mairi <http://orcid.org/orcid.org/0000-0002-9307-8565>



**Published version**

ALRENI, Ahmad Salah Eldin, HARROP, Deborah, LOWE, Anna, POTIA, Tanzila, KILNER, Karen and MCLEAN, Sionnadh Mairi (2017). Measures of upper limb function for people with neck pain. A systematic review of measurement and practical properties. Musculoskeletal Science and Practice, 29, 155-163.

# Measures of upper limb function for people with neck pain. A systematic review of measurement and practical properties

Ahmad Salah Eldin Alreni a, *
ahmadalreni@hotmail.co.uk
Deborah Harrop a
d.harrop@shu.ac.uk
Anna Lowe a
a.lowe@shu.ac.uk
Tanzila Potia a
tanzila.p@gmail.com
Karen Kilner a
k.kilner@shu.ac.uk
Sionnadh Mairi McLean b
s.mclean@shu.ac.uk
a Sheffield Hallam University, Montgomery House, 32 Collegiate Crescent, Collegiate Campus, Sheffield, S10 2BP, UK
b Sheffield Hallam University, Mercury House, 36 Collegiate Crescent, Collegiate Campus, Sheffield, S10 2BP, UK
*Corresponding author.

## Abstract

There is a strong relationship between neck pain (NP) and upper limb disability (ULD). Optimal management of NP should incorporate upper limb rehabilitation and therefore include the use of an ULD measure in the assessment and management process. Clear guidance regarding the suitability of available measures does not exist. The aim of this study was to identify all available measures of ULD for populations with NP, critically evaluate their measurement properties and finally recommend a list of suitable measures. This two-phase systematic review is reported in accordance with the PRISMA statement. Phase one identified clearly reproducible measures of ULD for patients with NP. Phase two identified evidence of their measurement properties. In total, 11 papers evaluating the measurement properties of five instruments were included in this review. The instruments identified were the DASH questionnaire, the QuickDASH questionnaire, the NULI questionnaire, the SFA and the SAMP test. There was limited positive evidence of validity of the DASH, QuickDASH, NULI, SFA and SAMP. There was limited positive evidence of reliability of the NULI, SFA and SAMP. There was unknown evidence of responsiveness of the DASH and QuickDASH. Although all measures are supported by a limited amount of low quality evidence, the DASH, QuickDASH, NULI questionnaires, and the SAMP test are promising measures, but they require further robust evaluation.

**Keywords**: Upper limb function; Disability; Neck pain; Outcome measures; Quality; Psychometric properties

## 1. Introduction

There is a strong relationship between neck pain and upper limb disability. The presence of a neck disorder is a risk factor for the development of an upper limb disability (Hakala et al.,2002; Walker-Bone et al., 2004; Frank et al., 2005; Bot et al., 2005; Huisstede et al., 2006; Rasmussen et al., 2008; Feleus et al., 2008). A clear example of this is cervical radiculopathy which can lead to pain, motor weakness, sensory deficit and loss of function in the neck, shoulder, upper arm or forearm (Polston, 2007; Rhee et al., 2007). Non-specific neck pain has also been shown to have a considerable impact on upper limb function (Frank et al., 2005; McLean et al., 2010a; Mclean et al., 2011; Osborn and Jull, 2013). In addition, coexisting shoulder dysfunction may also lead to neck pain becoming a recurrent, persistent or disabling problem (Eriksen et al., 1999; Bot et al., 2005; McLean et., 2010a). This can have a substantial effect on quality of life, work absenteeism, loss of productive capacity and consequently a substantial economic burden for patients, employers, insurers and society (Silverstein et al., 1998; Walker-Bone et al., 2002; Daffner et al., 2003; Baldwin and Butler, 2006; Cote et al., 2008).

The mechanisms which cause neck pain and upper limb disability to coexist are not clear, but may relate to the mechanical attachment between the neck and the upper limb via skeletal, muscular and neural structures (Mclean et al., 2011). For example, mechanical loading or repetitive movement of the upper limb may increase the mechanical load to the articular and ligamentous structures of the neck which may in turn provoke neck pain or create protective neck muscles spasm (Gorski and Schwartz, 2003). Another possible mechanism is that patients with neck pain may limit the functional use of their upper limb because of neck pain provocation or poor pain self-efficacy (Mclean et al., 2011). Consequently, a deconditioning effect may occur leading to a reduction in cardiovascular capacity and reduced strength and endurance in the neck/upper limb muscles. This altered upper limb conditioning may lead to compensatory activity and excessive loading on the cervical structures (Smeets et al., 2006). In the examination of neck pain, clinical textbooks frequently recommend simple screening of shoulder range of motion to rule in/out the presence of upper limb pain/disability (Petty, 2011). However, this may not be sufficient because range of motion does not conclusively correlate with disability (Poitras et al., 2000; Olson et al., 2000; Kwak et al., 2005). Optimal assessment of neck pain requires additional evaluation of the upper limb functional capacity and this suggests the utilisation of a suitable upper limb outcome measure (Mclean et al., 2011; Osborn and Jull, 2013). This would enable the accurate identification and quantification of any upper limb disability that may be present in a patient with neck pain and to evaluate the effectiveness of upper limb rehabilitation in the management plan (Connell and Tyson, 2012).

There is no clear guidance on the availability and suitability of instruments that measure upper limb functional capacity in patients with neck pain. Therefore, the aims of this review were to identify, summarise and critically examine all measures developed or evaluated to assess upper limb functional capacity in patients with neck pain and recommend relevant and suitable measures of upper limb function for patients with neck pain for application in clinical practice and research practice.

## 2. Methods

This systematic review was conducted in two phases. Phase one identified measures that have been used to assess upper limb functional capacity in patients with neck pain. Phase two identified studies evaluating the measurement and practical properties of the identified measures. The methodological quality of the developmental and/or evaluative studies of those identified measures were assessed against the "COnsensus-based Standards for the selection of health Measurement INstruments" (COSMIN) checklist (Mokkink et al., 2010; Terwee et al., 2012). The results were reported following the Preferred Reporting Items for Systematic Review and Meta-Analysis (PRISMA) statement (Liberati et al., 2009).

### 2.1 Phase one – identification of measures

The bibliographic databases as follows were searched from their inception to March 2016: Allied and Complementary Medicine Database (AMED) (OvidSP), CINAHL Complete (EBSCO), Cochrane Library (Wiley), MEDLINE (EBSCO), PubMed (US National Library of Medicine), PsycINFO (ProQuest), SPORTDiscus (EBSCO), Web of Science (Thomson Reuters).

#### 2.1.1 Search strategy

The search strategy comprised terms relating to upper limb function, and neck pain, and incorporated terms to limit to outcome measures, psychometric or measurement properties. The searches were undertaken in February and March 2016. All search terms were looked for in the title and abstract fields and controlled vocabulary terms were used where available. The Boolean operators AND and OR were used, alongside truncation, phrase searching and proximity operators. A copy of the search strategy for MEDLINE (EBSCO) is located in Appendix 1. The search syntax detailed in Appendix 1 were adapted for use on the other information resources used in the search.

### 2.1.2 Study selection

All articles yielded from the literature searches were eligible for inclusion in this review without restriction of study design or publication date provided the article: (1) was a full-text original primary quantitative study (e.g. clinical trials, observation studies, longitudinal studies, case controlled studies or case studies), (2) was published in the English language, (3) involved adults ≥ 18 years of age with neck pain (neck pain is defined here as dysfunction of the cervical structures (WHO, 2001), and (4) contained at least one measure of upper limb disability (upper limb disability is defined here as any difficulties or limitation an individual may have in executing upper limb activity (WHO, 2001). Articles were excluded if they did not use primary quantitative data, (e.g. systematic reviews, meta-analysis, qualitative studies, reportage or opinion pieces), the outcome measures did not measure upper limb disability in patients with neck pain, or involved participants with disorders other than neck pain.

Two reviewers (ASEA and AL) independently screened the title and abstract of all articles retrieved from the literature searches to determine their eligibility for inclusion in this review. This was followed by a full-text screening of all remaining articles to further determine their eligibility for inclusion in this review. In case of a disagreement between the two reviewers as to whether an article should be included or excluded, a consensus was sought through discussion, and if required a third reviewer (SMM) made the final decision. The reference lists of all included articles were screened by two reviewers (ASEA and AL) in order to identify additional relevant articles.

## 2.2 Phase two – identification of the development and/or evaluative studies

A second search was performed, using the databases identical to those searched in phase one. The name of each instrument identified in phase one was searched for using the all fields search function and was used to identify all articles related to the development or evaluation of the measurement properties of this instrument. A sensitive search filter, as reported by Terwee et al. (2009), was used to locate articles reporting the measurement properties of each identified instrument. Furthermore, the developers of specific measures were contacted to request additional evidence of measurement evaluation.

## 2.3 Data extraction

A data extraction form informed by earlier reviews from Haywood et al. (2013, 2014) and the COSMIN checklist (Mokkink et al., 2010; Terwee et al., 2012) was used to capture study specific (population, intervention, and setting) and

measurement specific information: reliability (internal consistency, test-retest, intra-/inter tester, measurement error), validity (face/content, structural validity (dimensionality), construct validity (evidence of explicit hypothesis testing, discriminant/discriminative), criterion validity (concurrent, predictive), responsiveness (criterion approach, construct approach), interpretability (for example, evidence of minimal important change), data precision (data quality, end effect), and evidence of where Item Response Theory (IRT) models where applied. Extraction of practical properties included acceptability (relevance and respondent burden) and feasibility (clinician burden, including cost, time to complete/score). The extent of patient involvement in measurement development and/or application was also sought (Haywood et al., 2014).

### 2.4 Quality assessment

In accordance with the COSMIN checklist, study methodological quality was evaluated for each measurement property investigated within the study and rated on a four-point scale (excellent, good, fair or poor) (Mokkink et al., 2010; Terwee et al., 2012). The quality rating of a study was determined by the lowest rating of any COSMIN checklist item related to the assessment of a specific measurement property i.e. "worst score counts" (Terwee et al., 2012). For example, the methodological quality of the study is considered excellent if all items related to a specific measurement property are rated as excellent. However, if any item is rated as poor, the methodological quality of the study is also rated as poor for that measurement property (Terwee et al., 2012).

Two reviewers (ASEA and TP), independently performed the data extraction and quality assessment of all included studies. In the case of disagreement about a study, a consensus was reached between the two reviewers via discussion. A third reviewer (SMM) was available to make the final decision if necessary.

### 2.5 Data analysis

Data was qualitatively synthesised using best evidence synthesis, to determine the overall quality and acceptability of each identified measure (Haywood et al., 2013, 2014). Different studies on the measurement properties of each identified measure were summarised by combining their result based on: (1) the number of studies in which the measurement property was assessed, (2) their methodological quality (COSMIN score) and (3) the consistency of the results of each measurement property. The overall rating of each measurement property was considered positive (+), negative (−), or indeterminate (?) following the criteria reported by (Terwee et al., 2007) Table 1. This was accompanied by the level of evidence suggested by the Cochrane Back Review Group in which the possible level of evidence for a measurement property is

“strong”, “moderate”, “limited”, “conflicting” or “unknown” (van Tulder et al., 2003; Furlan et al., 2009) Table 2.

**Table 1** Quality criteria for measurement properties. (Terwee et al., 2007).

| Property | Rating† | Quality Criteria |
|---|---|---|
| **Reliability** | | |
| Internal consistency | + | (Sub)scale unidimensional AND Cronbach's alpha(s) ≥ 0.70 |
| | ? | Dimensionality not known OR Cronbach's alpha not determined |
| | - | (Sub)scale not unidimensional OR Cronbach's alpha(s) < 0.70 |
| Reliability | + | ICC / weighted Kappa ≥ 0.70 OR Pearson's r ≥ 0.80 |
| | ? | Neither ICC / weighted Kappa, nor Pearson's r determined |
| | - | ICC / weighted Kappa < 0.70 OR Pearson's r < 0.80 |
| Measurement error | + | MIC > SDC OR MIC outside the LOA |
| | ? | MIC not defined |
| | - | MIC ≤ SDC OR MIC equals or inside LOA |
| **Validity** | | |
| Content validity | + | All items are considered to be relevant for the construct to be measured, for the target population, and for the purpose of the measurement AND the questionnaire is considered to be comprehensive |
| | ? | Not enough information available OR no target population invovement |
| | - | Not all items are considered to be relevant for the construct to be measured, for the target population, and for the purpose of the measurement OR the questionnaire is considered not to be comprehensive |
| Construct validity - Structural validity | + | Factors should explain at least 50% of the variance |
| | ? | Explained variance not mentioned |
| | - | Factors explain < 50% of the variance |
| - Hypothesis testing | + | Correlations with instruments measuring the same construct ≥ 0.50 OR at least 75% of the results are in accordance with the hypotheses AND correlations with related constructs are higher than with unrelated constructs |
| | ? | Solely correlations determined with unrelated constructs |
| | - | Correlations with instruments measuring the same construct < 0.50 OR < 75% of the results are in accordance with the hypotheses OR correlations with related constructs are lower than with unrelated constructs |
| - Cross-cultural validity | + | No differences in factor structure OR no important DIF between language versions |
| | ? | Multiple group factor analysis not applied AND DIF not assessed |
| | - | Differences in factor structure OR important DIF between language versions |
| Criterion validity | + | Convincing arguments that gold standard is "gold" AND correlation with gold standard ≥ 0.70 |
| | ? | No convincing arguments that gold standard is "gold" |
| | - | Correlation with gold standard < 0.70 |
| **Responsiveness** | | |
| Responsiveness | + | Correlation with changes on instruments measuring the same construct ≥ 0.50 OR at least 75% of the results are in accordance with the hypotheses OR AUC ≥ 0.70 AND correlations with changes in related constructs are higher than with unrelated constructs |
| | ? | Solely correlations determined with unrelated constructs |
| | - | Correlations with changes on instruments measuring the same construct < 0.50 OR < 75% of the results are in accordance with the hypotheses OR AUC < 0.70 OR correlations with changes in related constructs are lower than with unrelated constructs |

[..] Reference Number, MIC Minimal Important Change, SDC Smallest Detectable Change, LOA Limits of Agreement, ICC Interclass Correlation Coefficient, AUC Area Under the Curve.
+ = positive rating, - = negative rating, ? = indeterminate rating.

**Table 2** Level of evidence for the overall quality of measurement property.

| Level | Rating[†] | Criteria |
|---|---|---|
| strong | **+++ or --** | Consistent findings in multiple studies of good methodological quality OR in one study of excellent methodological quality |
| moderate | **++ or --** | Consistent findings in multiple studies of fair methodological quality OR in one study of good methodological quality |
| limited | **+ or -** | One study of fair methodological quality |
| conflicting | **+/-** | Conflicting findings |
| unknown | **?** | Only studies of poor methodological quality |

+ = positive result, − = negative result. (van Tulder et al., 2003; Furlan et al., 2009).

## 3. Results

### 3.1 Identification of studies and measures

#### 3.1.1 Phase one

Following the removal of duplicates, 982 unique records were identified from the database searches. Following a title and abstract screening process, 54 articles were retained. Following a full text reading of the remaining 54 articles, five articles were retained for inclusion in this review. Screening of the reference lists from included articles resulted in 15 potentially relevant articles, of which one article met the inclusion criteria for this review. Only five clearly described and reproducible instruments were included.

#### 3.1.2 Phase two

Evidence for the measurement and/or practical properties were sought for the five reproducible measures identified in phase one. However, the database searches did not uncover any new records. Contacting the developers of specific measures resulted in six additional articles, of which five were retained for inclusion in the review.

#### 3.1.3 Results from phase one and phase two

In total, 11 articles on the development/evaluation of five measures were included in this review. Fig. 1 shows the phase one and phase 2 outcomes at each stage of the selection/screening process and the reasons for exclusions.

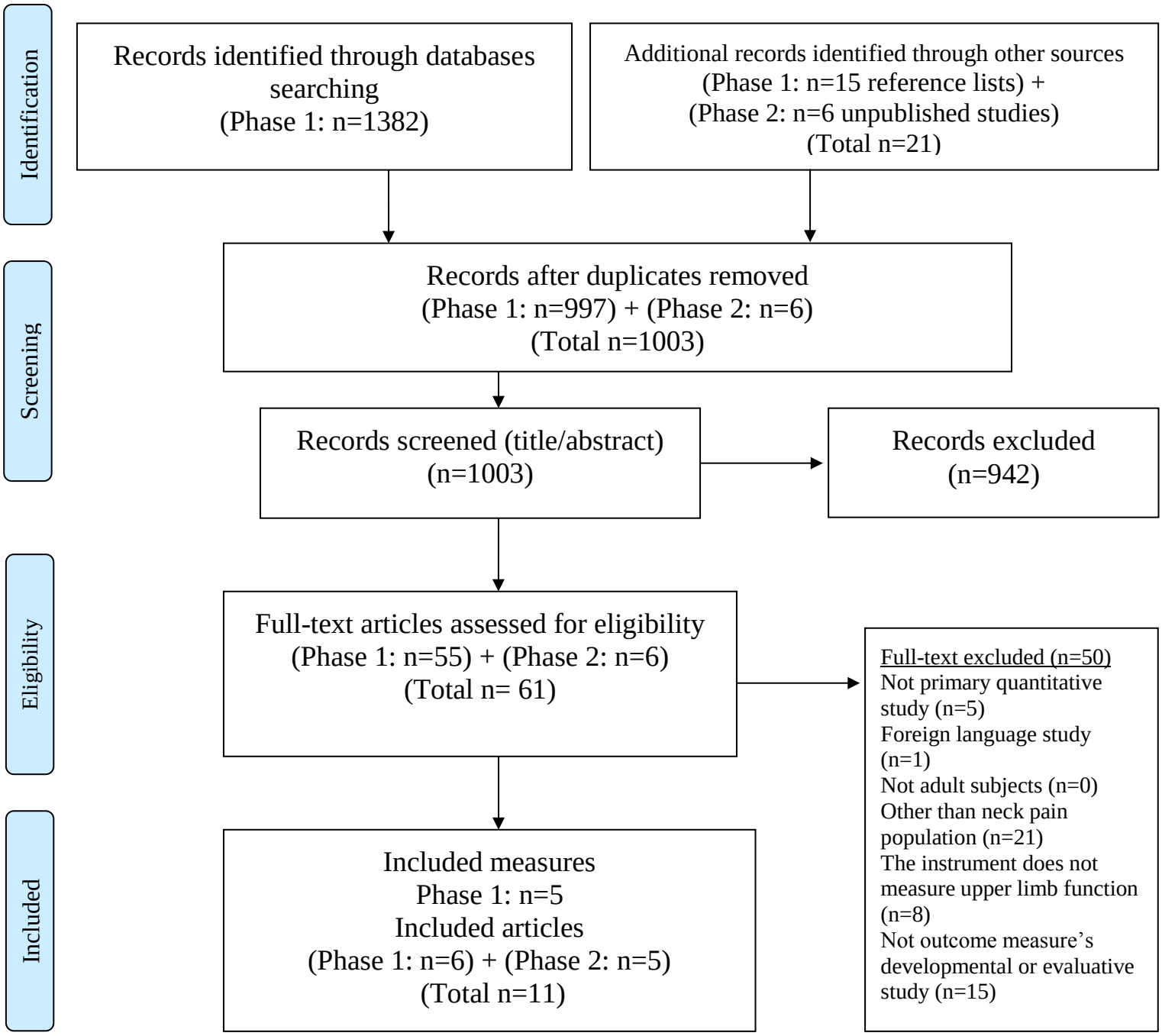


**Fig. 1.** PRISMA flow chart of phase one and phase two.

These 11 articles provide evidence for five clearly defined and reproducible measures of upper limb disability in the context of neck pain. Three are patient-reported: The Disability of Arm, Shoulder, and Hand (DASH) questionnaire (Hudak et al., 1996), the Quick Disability of Arm, Shoulder and Hand (QuickDASH) questionnaire (Beaton et al., 2005) and the Neck and Upper Limb Index (NULI) questionnaire (Stock et al., 2003). One is clinician reported: The Shoulder Functional Assessment (SFA) (Lomond and Cote, 2009). One is performance-based: The Single Arm Military Press (SAMP) test (McLean et al., 2010b). The general characteristics of the 11 articles are presented in Table 3. The methodological quality of each study per measurement property is presented in Table 4. A synthesis of the results for each instrument a long with their level of evidence is presented in Table 5. Questionnaires such as the Rapid Upper Limb Assessment (RULA) (McAtammy and Corlett, 1993) and the Maastricht Upper Extremity Questionnaire (MUEQ) (Eltayeb et al., 2007) were excluded since they are risk assessment questionnaires rather than outcome measures. A summary of measurement properties for each identified instrument follows.

**Table 3** Characteristics of the included studies.

| Study | Sample Size | Mean age ± SD/range | Population | Country | Setting | Recruitment methods | Outcome measures used in the study | Measurement property assessed |
|---|---|---|---|---|---|---|---|---|
| Huisstede et al. (2009) | N=679 | 44.4 ± 11.4 (18-64) | Neck, shoulder, and/or arm pain | Holland | Dutch General Practices (GPs) | Convenience | DASH<br>SF-12<br>Severity of complaint<br>Persistence of complaint | Hypothesis testing<br>Responsiveness |
| Mehta et al. (2010) | N=66 | 40.6 ± 14.2 | Neck pain with/without arm pain, headache and whiplash disorders level 2&3 | CANADA | Canadian Physical Therapy Clinics | Convenience | DASH<br>QDASH<br>NDI<br>VAS<br>CSOQ | Hypothesis testing<br>Concurrent validity |
| Fan et al. (2008) | N=733 | N=733 39.5 ± 0.05<br>N=321 43.2 ± 0.7<br>N=175 39.3 ± 0.8 | Neck Or Upper Extremity Musculoskeletal Disorders (UEMSDs) | USA | Workplace walkthrough at 12 manufacturing and service work sites in Washington State | Convenience | QDASH<br>SF-12<br>Symptoms severity | Hypothesis testing<br>Concurrent validity<br>Predictive validity |
| Fan et al. (2011) | N= 465 | N=50 35.3 ± 10.2<br>N=18 42.6 ± 10.9<br>N=46 35.5 ± 10.2<br>N=34 41.9 ± 11.3<br>N=317 41.1 ± 10.7 | Neck Or Upper Extremity Musculoskeletal Disorders (UEMSDs) | USA | Workplace walkthrough at 12 manufacturing and service work sites in | Convenience | QDASH<br>SF-12<br>QDASH work module<br>Severity | Responsiveness |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | | | | | Washington State | | | |
| Stock et al. (2003) | Ontario N=119 Quebec N=93 | Ontario: 39.7 ± 10.1 Quebec: 41.1 ± 10.0 | Workers with neck and upper limb dysfunction | CANADA | Workers from community private physiotherapy clinics | Convenience | NULI<br>SIP<br>SF-36 | Internal consistency<br>Reliability<br>Structural validity<br>Hypothesis testing<br>Responsiveness |
| Lomond and Cote (2009) | N=32 | N=16 40.1 ± 12.1<br>N=16 39.7 ± 13.2 | Chronic neck and shoulder pain | CANADA | Institutional rehabilitation programme, advertisement, research centre staff and social network | Convenience | SFA<br>SPADI<br>NDI<br>NRS<br>The Borg CR-10 scale | Test-retest, inter, intra-rater reliability<br>Measurement error<br>Hypothesis testing |
| Patekar (2010) | N=98 | 42.2 ± 7.85 (30-60) | Non-patients subjects with and without neck symptoms | UK | Institutional staff and students (institutional campus) | Convenience | SAMP | Hypothesis testing |
| Darne (2010) | N=95 | 44.53 ± 7.9 (30-60) | Non-patient subjects with and without neck symptoms | UK | Institutional staff and students (institutional campus) | Convenience | SAMP<br>DASH | Hypothesis testing |
| Toulassidharane (2010) | N=190 | 41.8 ± 8.1 (30-59) | Non-patient subjects with and without neck symptoms | UK | Institutional staff and students (institutional campus) | Convenience | SAMP<br>DASH | Hypothesis testing |
| Kulkarni (2010) | N=95 | 38.95 ± 7.22 (30-60) | Non-patient subjects with and without neck symptoms | UK | Institutional staff and students (institution | Convenience | SAMP<br>DASH | Test-retest, inter, intra-rater reliability |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | | | | | al campus) | | | |
| Jain (2010) | N=95 | 44.5 ± 7.9 (30-60 | Non-patient subjects with and without neck symptoms | UK | Institutional staff and students (institutional campus) | Convenience | SAMP DASH | Test-retest, inter, intra-rater reliability |

**Table 4** Methodological qualities of each study per measurement property.

| Study | Internal consistency **(A)** | Test-retest, inter, intra-rater reliability **(B)** | Measurement error **(C)** | Content validity **(D)** | Structural validity **(E)** | Hypothesis testing **(F)** | Concurrent validity **(H)** | Predictive validity **(H)** | Responsiveness **(I)** |
|---|---|---|---|---|---|---|---|---|---|
| **DASH** | | | | | | | | | |
| Huisstede et al. (2009) | | | | | | Poor | | | Poor |
| Mehta et al. (2010) | | | | | | Poor | Poor | | |
| **QDASH** | | | | | | | | | |
| Fan et al. (2008) | | | | | | Poor | Poor | Poor | |
| Fan et al. (2011) | | | | | | | | | Poor |
| Mehta et al. (2010) | | | | | | Poor | Poor | | |
| NULI | | | | | | | | | |
| Stock et al. (2003) | Fair | Fair | | | Fair | Fair | | | Poor |
| **SFA** | | | | | | | | | |
| Lomond and Cote. (2009) | | Fair | Fair | | | Poor | | | |
| **SAMP** | | | | | | | | | |
| Patekar (2010) | | | | | | Fair | | | |
| Darne (2010) | | | | | | Poor | | | |
| Kulkarni (2010) | | Fair | | | | | | | |
| Toulassidharane (2010) | | | | | | Poor | | | |
| Jain (2010) | | Fair | | | | | | | |

(..) reference number, (DASH) Disability of Arm, Shoulder and Hand, (QDASH) Quick Disability of Arm, Shoulder and Hand, (SAMP) Single Arm Military Press. (NULI) Neck and Upper Limb Index. Study is mentioned twice because of evaluating measurement properties of two instruments.

**Table 5** Quality of measurement properties per instrument.

| Instrument | Internal consistency | Reliability | Measurement error | Content validity | Structural validity | Hypothesis testing | Criterion validity | Responsiveness | Practical properties | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | | | Precision | Acceptability | Feasibility |
| **DASH** | *na* | *na* | *na* | *na* | *na* | ? | ? | ? | *na* | *na* | *na* |
| | | | | | | | | | | | |
| **QDASH** | *na* | *na* | *na* | *na* | *na* | ? | ? | ? | *na* | *na* | *na* |
| | | | | | | | | | | | |
| **NULI** | + | + | *na* | *na* | + | + | *na* | ? | *na* | *na* | *na* |
| | | | | | | | | | | | |
| **SFA** | *na* | + | + | *na* | *na* | ? | *na* | *na* | *na* | *na* | *na* |
| | | | | | | | | | | | |
| **SAMP test** | *na* | ++ | *na* | *na* | *na* | + | *na* | *na* | *na* | *na* | *na* |

+++ or −−− strong evidence positive/negative results, ++ or − moderate evidence positive/negative results, + or − limited evidence positive/negative results, ± conflicting evidence, ? Unknown, due to poor methodological quality, na no information available.

### 3.2 DASH

The DASH is a patient-reported outcome (PRO) instrument developed to measure upper limb (hand, wrist, elbow and shoulder) disability/symptoms as a single functional unit (Hudak et al., 1996). The instrument uses 30 items, each scored on a 1–5 scale. A total score is calculated by summing item scores and transforming into a score from 0 to 100 where 0 equals no disability and 100 equals the most severe disability (Hudak et al., 1996). The DASH has been evaluated for use in the population with neck pain in two studies (Huisstede et al., 2009; Mehta et al., 2010).

There was no evidence identified for internal consistency, reliability, measurement error, face/content validity, structural validity or any of the practical properties for the DASH on the population with neck pain. There was also no evidence identified for patient involvement in the development of the questionnaire. There was unknown evidence for hypothesis testing of the DASH (Huisstede et al., 2009; Mehta et al., 2010). There was unknown evidence for responsiveness of the DASH (Huisstede et al., 2009); the use of Guyatt's responsive ratio is an inappropriate parameter of responsiveness (de Vet et al., 2011). There was unknown evidence for criterion validity of the DASH (Mehta et al., 2010) since the criterion employed cannot be considered an adequate 'gold standard' (de Vet et al., 2011).

### 3.3 QuickDASH

The QuickDASH is an 11-item questionnaire derived from the DASH and designed to be shorter measure of physical function and symptoms related to the upper-limb musculoskeletal disorders (Beaton et al., 2005). Each item is scored on a 1–5 scale and the total score is derived by summing item scores and transforming them into a score from 0 to 100, where 0 equals no disability and 100 equals the most severe disability. The QuickDASH has been evaluated for use in populations with neck pain in three studies (Fan et al., 2008, 2011; Mehta et al., 2010).

There was no evidence identified for internal consistency, reliability, measurement error, face/content validity, structural validity or any of the practical properties for the QuickDASH in populations with neck pain. There was also no evidence identified of patient involvement in the development of the questionnaire. There was unknown evidence for hypothesis testing of the QuickDASH (Mehta et al., 2010). There was unknown evidence for criterion validity since the criterion used (Neck Disability Index) cannot be considered an adequate 'gold standard' (Fan et al., 2008; Mehta et al., 2010). There was unknown evidence of responsiveness (Fan et al., 2011); the use of Effect Size

(ES) and Standardised Response Mean (SRM) are inappropriate parameters of responsiveness (de Vet et al., 2011).

### 3.4 NULI

The NULI is a short English and French language patient-reported measure of musculoskeletal dysfunction of the neck and upper limb for Canadian workers (Stock et al., 1995; Stock 2000; Salerno et al., 2002; Stock et al., 2003). It is a 20-item multidimensional index evaluating the impact of neck and upper limb dysfunction on physical activity (7 items), work related (4 items), psychosocial (6 items), sleep related (2 items) and 1 item related to the iatrogenic effect of assessment and treatment (Stock et al., 2003). Section A, questions 1–11 are scored on a 1–7 scale, where 1 equals no difficulties at all and 7 equals cannot do. Section B, questions 12–20 are scored on a 1–7 scale where 1 equals never and 7 equals all the time (Stock et al., 2003).

The NULI was developed and evaluated in one study (Stock et al., 2003). There was no evidence for measurement error, face/content validity, criterion validity, score interpretation or any of practical properties for the NULI.
There was also no evidence identified for patient involvement in the development of the measure. There was limited positive evidence for internal consistency (Cronbach Alpha = 0.90 and 0.92) of the NULI for participants speaking English and French respectively (Stock et al., 2003). There was limited positive evidence for reliability of the NULI (ICC = 0.88 and 0.83) for participants speaking English and French respectively (Stock et al., 2003). There was limited positive evidence for structural validity of the NULI (Factor analysis used and the 20 items distributed well according to four dimensions) (Stock et al., 2003). There was limited positive evidence of hypothesis testing (convergent validity) (r = 0.3–0.5, 0.4–0.6, 0.6–0.73, 0.75) for physical activity, pain, work and psychosocial dimensions respectively (Stock et al., 2003). There was unknown evidence of responsiveness of the NULI (Stock et al., 2003).

### 3.5 SFA

The SFA is a clinician-reported measure developed to measure shoulder functional capacity in workers with chronic neck/shoulder pain and healthy subjects (Lomond and Cote, 2009). It involves a series of shoulder functional tasks such as active shoulder range of motion in flexion/abduction and repetitive pushing/pulling utilising the Baltimore Therapeutic Equipment Work Simulator II (Sim II) (BTE-Tech©, Baltimore, MD) (Lomond and Cote, 2009).

There was no evidence identified for face/content validity, criterion validity, responsiveness, score interpretation or any practical properties for the SFA.

There was also no evidence of patient involvement in the development of the measure. There was limited positive evidence for reliability of the SFA (ICC Flexion Range of Motion (ROM), Abduction (ROM), Cumulative Power Output = 0.95–0.92, 0.85–0.87, 0.94–0.53 respectively) for control and pain groups respectively (Lomond and Cote, 2009). There was limited positive evidence for measurement error of the SFA (SEM Flexion (ROM), Abduction (ROM), Cumulative Power Output = 4.72–14.76, 6.06–24.35, 7.52–30.25) for control and pain groups respectively (Lomond and Cote, 2009). There was unknown evidence for hypothesis testing of the SFA (Lomond and Cote, 2009).

### 3.6 SAMP test

The SAMP test is a performance-based test developed to measure upper limb functional capacity and it was specifically developed for use in populations with neck pain (McLean et al., 2010b; Pateker 2010; Darne 2010; Kulkarni 2010; Toulassidharane 2010; Jain 2010). The test involves repeatedly lifting a 3kg hand-weight overhead from the shoulder level for 30 s. The SAMP score is the number of repetitions correctly completed; higher scores represent a lower level of upper limb disability (McLean et al., 2010b).

There was no evidence identified for measurement error, face/content validity, criterion validity, responsiveness or any practical properties for the SAMP test. There was also no evidence of patient involvement in the development of the measure. There was moderate positive evidence for reliability of the SAMP test (ICC = 0.94–0.99 and 0.982–0.977) for asymptomatic and symptomatic participants respectively (Jain 2010; Kulkarni 2010). There was limited positive evidence for hypothesis testing of the SAMP test (r = 0.814) (Patekar 2010).

## 4. Discussion

This review identified five measures used to evaluate upper limb disability in populations with neck pain and 11 studies evaluating their measurement properties. Significant methodological and quality issues prevent a clear recommendation for any of the identified measures. Evidence for the five identified and reviewed measures was limited, unknown or unavailable. Only one measure is performance-based, the SAMP test, that was developed specifically for use in populations with neck pain.

There is substantial evidence that the DASH and QuickDASH are strongly performing measures (Bot et al., 2004; Huang et al., 2015; Kennedy et al., 2013) and limited evidence that the NULI and SFA are reliable and valid measures (Lomond and Cote, 2009; Stock et al., 2003) in population with shoulder or upper limb problems. However, application of a measure which is

inadequately developed/evaluated, or for a purpose other than which it was intended threatens its validity and limits meaningful interpretation with which to inform decision-making regarding the management plan.

Evidence of face/content validity and practical properties including acceptability and feasibility was not identified for any of the reviewed measures. Acceptability, appropriateness and feasibility of measures for patients and clinicians, considered vital for the utility of a measure (Tyson et al., 2008; Connell and Tyson, 2012), were not considered in any of the studies. There was no evidence of patient involvement in the development/evaluation of any of the measures. Patient involvement as a research partner is considered essential to ensure the relevance and validity of patient-centred outcome assessment (Mayer, 2012; Staniszewska et al., 2012). One clinician-reported measure (SFA) was developed for workers only and it involves the use of very expensive equipment, which is likely to limits its use in clinical practice (Tyson et al., 2008). The QuickDASH, NULI and SAMP test are quick, economical and easy to administer and score. Since brevity is crucial in clinical practice, QuickDASH, NULI and the SAMP test are considered to be promising measures of upper limb disability for the population with neck pain, however, further adequate evaluation is strongly recommended.

The strength of this review relates to the transparent evaluation of the identified studies and measurement quality using the COSMIN checklist (Mokkink et al., 2010; Terwee et al., 2012), and to the reporting of results in accordance with the PRISMA statement (Liberati et al., 2009). In addition, this is the first systematic review that has sought to identify and evaluate the measurement properties of all available measures of upper limb function developed or evaluated for use in the population with neck pain. Comprehensive search strategies that utilised multiple major databases and involved multiple authors in the development, review and refinement were used in this review. Although these search strategies were limited to English-language publications, English-language abstracts for non-English publication were reviewed and one study only was excluded, and this was due to irrelevance not language. This suggests that the likelihood of selection bias is low.

The level of evidence criteria in Table 2 which was suggested by the Cochrane Back Review Group (van Tulder et al., 2003; Furlan et al., 2009) was originally proposed for systematic reviews conducted on clinical trials. However, it has been used in similar studies and found to be applicable to reviews investigating the measurement properties of health-related outcome measures (Schellingerhout et al., 2011, 2012).

## 5. Conclusion

In the absence of high quality studies and inadequate reporting of essential measurement and practical properties, application of the identified measures of upper limb disability should be undertaken cautiously in the population with neck pain until acceptable evidence is established. Further research should incorporate COSMIN recommendations during the design of developmental or evaluative studies of these measurement instruments. The involvement of key stakeholders, including patients and clinicians is essential to ensure that the measure is relevant, acceptable and feasible.

## Appendix 1. Search strategy

The search strategy has been written up for MEDLINE using the EBSCO interface and is detailed below.

Explanation of search terms used: ti = title field; ab = abstract field;/= MeSH; asterisk (*) denotes any character; "" = phrase search; N5 = adjacency within five words.

“upper limb”[ti,ab] OR “upper extremity”[ti,ab] OR function*[ti,ab] OR dysfunction*[ti,ab] OR abilit*[ti,ab] OR disabilit*[ti,ab] OR capacity*[ti,ab] OR disorder*[ti,ab] OR problem*[ti,ab] OR pain*[ti,ab] OR deficit*[ti,ab] AND neck[ti,ab] OR “cervical spine”[ti,ab] OR cervicogenic*[ti,ab] OR pain*[ti,ab] OR function*[ti,ab] OR dysfunction*[ti,ab] OR abilit*[ti,ab] OR disabilit*[ti,ab] OR problem*[ti,ab] OR disc*[ti,ab] OR “degenerative disc”[ti,ab] OR degeneration*[ti,ab] OR disease*[ti,ab] OR disorder*[ti,ab] OR deficit*[ti,ab] AND “outcome measure*” n5[ti,ab] OR “outcome assessment*”[ti,ab] OR psychometr*[ti,ab] OR clinimetr* [ti,ab] OR “observer variation*”[ti,ab] OR reproducib*[ti,ab] OR reliab*[ti,ab] OR unreliab*[ti,ab] OR valid*[ti,ab] OR discriminant*[ti,ab] OR coefficient*[ti,ab] OR correlation*[ti,ab] OR selection*[ti,ab] OR reduction* [ti,ab] OR agreement*[ti,ab] OR precision*[ti,ab] OR imprecision*[ti,ab] OR test retest*[ti,ab] OR interrater*[ti,ab] OR intrarater*[ti,ab] OR inter-rater*[ti,ab] OR intra-rater*[ti,ab] OR kappa*[ti,ab] OR “minimal important change*”[ti,ab] OR “multitrait scaling analysis*”[ti,ab] OR “factor analysis*”[ti,ab] OR “known group*”[ti,ab] OR responsive*[ti,ab].

## References


Baldwin ML. and Butler RJ. Upper extremity disorders in the workplace: Costs and outcomes beyond the first return to work. Journal of Occupational Rehabilitation 2006; 16:303–323.

Beaton DE, Wright JG, Katz JN. Development of the QuickDASH: comparison of three item-reduction approaches. J Bone Joint Surg Am 2005; 87:1038–46.

Bot S, Terwee CB, van der Windt DAWM, Bouter LM, Dekker J, de Vet HCW. Clinimetric evaluation of shoulder disability questionnaires: a systematic review of the literature. Ann Rheum Dis 2004; 63:335-341.

Bot S, van der Waal J, Terwee C, van der Windt D, Scholten R, Bouter L, et al. Predictors of outcome in neck and shoulder symptoms: a cohort study in general practice. Spine 2005; 30:459–70.

Connell L, Tyson S. Clinical reality of measuring upper-limb ability in neurologic conditions: a systematic review. Arch Phys Med Rehabil 2012; 93:221–8.

Cote P, van der Velde G, Cassidy JD, et al. The burden and determinants of neck pain in workers: results of the Bone and Joint Decade 2000–2010 Task Force on Neck Pain and its associated disorders. Spine 2008; 33:60–74.

Daffner SDM, Hilibrand AS, Hanscom BS, Brislin BT, Vaccaro AR, Albert TJ. Impact of neck and arm pain on overall health status. Spine 2003; 28:2030–5.

Darne R. Construct validity of the single arm military press (SAMP) test in females with neck pain. Sheffield Hallam University 2010.

de Vet HCW, Terwee CB, Mokkink LB, Knol DL. Measurement in Medicine: A Practical Guide to Biostatistics and Epidemiology. London: Cambridge University Press; 2011.

Eriksen W, Natvig B, Knardahl S, Bruusgaard D. Job characteristics as predictors of neck pain. A 4-year prospective study. JOEM/ACOEM 1999; 41:893-902.

Eltayeb S, Staal J, Kennes J, Lamberts P, Bie R. Prevalence of complaints of arm, neck and shoulder among computer office workers and psychometric evaluation of a risk factor questionnaire. BMC 2007; 8:68.

Fan ZJ, Smith CK, Silverstein BA. Assessing validity of the QuickDASH and SF-12 as surveillance tools among workers with neck or upper extremity musculoskeletal disorders. J Hand Ther 2008; 21:354–65.

Fan ZJ, Smith CK, Silverstein BA. Responsiveness of the QuickDASH and SF-12 in workers with neck or upper extremity musculoskeletal disorders: one-year follow-up. J Occup Rehabil. 2011; 21:234-243.

Feleus A, Bierma-Zeinstra SM, Miedema HS, et al. Incidence of Nontraumatic complaints of arm, neck and shoulder in general practice. Man Ther 2008; 13:426–33.

Frank A, De Souza L, Frank C. Neck pain and disability: a cross-sectional survey of the demographic and clinical characteristics of neck pain seen in a rheumatology clinic. Int J Clin Pract 2005; 59:173–82.

Furlan AD, Pennick V, Bombardier C, van Tulder M. Editorial Board, Cochrane Back Review Group. 2009 updated method guidelines for systematic reviews in the Cochrane Back Review Group. Spine 2009; 34:1929–41.

Gorski JM, Schwartz LH. Shoulder impingement presenting as neck pain. The Journal of Bone and Joint Surgery (American Volume) 2003; 85:635-8.

Hakala P, Rimpela A, Salminen JJ, et al. Back, neck, and shoulder pain in Finnish adolescents: national cross sectional surveys BMJ 2002; 325:743.

Haywood KLBJ, Tutton E, Staniszewska S. Patient-reported outcome measures (PROMs) and fragility hip fracture: a systematic review of quality and acceptability. Qual Life Res 2013; 22 Suppl 1:2077.

Haywood KL, Collin SM, Crawley E. Assessing severity of illness and outcomes of treatment in children with chronic fatigue syndrome/myalgic encephalomyelitis (CFS/ME): a systematic review of patient-reported outcome measures (PROMs). Child Care Health Dev 2014; 40:806–24.

Huang h, Grant JA, Miller BS, Mirza FM, Gagnier JJ. A systematic review of the psychometric properties of patient-reported outcome instruments for use in patients with rotator cuff disease. The American Journal of Sports Medicine 2015; 43:2572-2582.

Hudak PL, Amadio PC, Bombardier C; The Upper Extremity Collaborative Group (UECG). Development of an upper extremity outcome measure: the DASH (disabilities of the arm, shoulder and hand). AmJ Ind Med 1996;

29:602–8.

Huisstede B, Bierma-Zeinstra S, Koes BW, Verhaar JA. Incidence and prevalence of upper extremity musculoskeletal disorders: a systematic appraisal of the literature. BMC Musculoskelet Disord 2006; 7:7.

Huisstede B, Feleus A, Bierma-Zeinstra S, Verhaar JA, Koes BW. Is the disability of arm, shoulder, and hand questionnaire (DASH) also valid and responsive in patients with neck complaints. Spine 2009; 34:130–8.

Jain V. Inter-rater and intra-rater reliability of the single arm military press (SAMP) test. Sheffield Hallam University. 2010.

Kennedy CA, Beaton DE, Smith P, et al. Measurement properties of the QuickDASH (Disabilities of the Arm, Shoulder and Hand) outcome measure and cross-culture adaptations of the QuickDASH: a systematic review. Quality of Life Research 2013; 22:2509-2547.

Kulkarni M. Inter-rater and intra-rater reliability for the single arm military press (SAMP) test. Sheffield Hallam University. 2010.

Kwak S, Niederklein R, Tarcha R, Hughes C. Relationship between active cervical range of motion and perceived neck disability in community dwelling elderly individuals. Journal of Geriatric Physical Therapy 2005; 28:54-6.

Liberati A, Altman DG, Tetzlaff J, Mulrow C, Gotzsche PC, Ioannidis JP, et al. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration. Ann Intern Med 2009; 151:W65-94.

Lomond K. and Cote J. Shoulder functional assessments in persons with chronic neck/shoulder pain and healthy subjects: Reliability and effects of movement repetition. Work 2009; 38:169-180.

Mayer MM. Seeking what matters. The Patient-Patient-Centered Outcomes Research 2012; 5:71-74.

McAtammy L. and Corlett E. RULA: A survey method for the investigation of work-related upper limb disorders. Applied Ergonomics 1993; 24:91-99.

Mclean SM, Moffett J, Sharp D, Gardiner E. An investigation to determine the association between neck pain and upper limb disability for patients with non-specific neck pain: a secondary analysis. Man Ther 2011; 16:434–9.

McLean SM, May S, Moffett J, Sharp D, Gardiner E. Risk factors for the onset of non-specific neck pain: a systematic review. J Epidemiol Community Health 2010a; 64:565-572.

McLean S, Taylor J, Balassoubramanien T, Kulkarni M, Patekar P, Darne R, et al. Measuring upper limb disability in non-specific neck pain: a clinical performance measure. International Journal of Physical Therapy and Rehabilitation 2010b; 1:44-52.

Mehta S, MacDermid J, Carlesso L, McPhee C. Concurrent validation of the DASH and the QuickDASH in comparison to neck-specific scales in patients with neck pain. Spine 2010; 35:2150-2156.

Mokkink LB, Terwee CB, Patrick DL, Alonso J, Stratford PW, Knol DL, et al. The COSMIN checklist for assessing the methodological quality of studies on measurement properties of health status measurement instruments: an international Delphi study. Qual Life Res 2010; 19:539–49.

Olson SL, O'Connor DP, Birmingham G, Broman P, Herrera L. Tender point sensitivity, range of motion, and perceived disability in subjects with neck pain. The Journal of Orthopaedic and Sports Physical Therapy 2000; 30:13-20.

Osborn W, Jull G. Patients with non-specific neck disorders commonly report upper limb disability. Man Ther 2013; 18:492–7.

Patekar P. Clinical utility of single arm military press (SAMP) test in females with neck pain. Sheffield Hallam University 2010.

Petty N. Neuromusculoskeletal examination and assessment: a handbook for therapists. *ELSEVIER Edinburgh*; 4th edition 2011.

Poitras S, Loisel P, Prince F, Lemaire J. Disability measurement in persons with back pain: a validity study of spinal range of motion and velocity. Archives of Physical Medicine and Rehabilitation 2000; 81:1394-400.

Polston DW. Cervical radiculopathy. Neurol Clin 2007; 25:373–85.

Rasmussen C, Leboeuf-Yde C, Hestbaek L, et al. Poor outcome in patients with spine-related leg or arm pain who are involved in compensation claims: a prospective study of patients in the secondary care sector. Scand J Rheumatol 2008; 37:462–8.

Rasmussen C, Leboeuf-Yde C, Hestbaek L, et al. Poor outcome in patients with

spine-related leg or arm pain who are involved in compensation claims: a prospective study of patients in the secondary care sector. Scand J Rheumatol 2008; 37:462–8.

Rhee JM, Yoon T, Riew KD. Cervical radiculopathy. J Am Acad Orthop Surg 2007; 15:486–94.

Salerno DF, Copley-Merriman C, Tayler TN, Shinogle J, Schulz RM. A review of functional status measures for workers with upper extremity disorders 2002; 59: 664-670.

Schellingerhout JM, Verhagen AP, Heymans MW, Koes BW, de Vat HC, Terwee CB. Measurement properties of disease-specific questionnaires in patients with neck pain: a systematic review. Qual Life Res 2012; 21:659-670.

Schellingerhout JM, Heymans MW, Verhagen AP, de Vat HC, Koes BW, Terwee CB. Measurement properties of translated versions of neck-specific questionnaires: a systematic review. BMC Medical Research Methodology 2011; 11:87.

Silverstein B, Welp E, Nelson N, Kalat J. Claims incidence of work-related disorders of the upper extremities: Washington State, 1987 through 1995. Am J Public Health 1998; 88:1827–33.

Smeets RJ, Wade D, Hidding A, Van Leeuwen PJ, Vlaeyen JW, Knottnerus JA. The association of physical deconditioning and chronic low back pain: a hypothesisoriented systematic review. Disability and Rehabilitation 2006; 28:673-93.

Staniszewska S, Haywood KL, Brett J, Tutton L. Patient and public involvement in patient-reported outcome measures. The Patient-Patient-Centered Outcomes Research 2012; 5:79-87.

Stock SR, Loisel P, Durand M, et al. L'indice d'impact de la douleur au cou et aux membres supérieurs sur la vie quotidienne (NULI: neck and upper limb index). Études et Recherches. 2003; R-355. http://www.irsst.qc.ca/media/documents/pubirsst/R-355.pdf.

Stock SR, Streiner D, Reardon R, et al. The impact of neck and upper limb musculoskeletal disorders on the lives of affected workers: development of a new functional status index. Qual Life Res 1995; 4:491.

Stock, SR. A brief overview of the Neck and Upper Limb Index (NULI).

Physical Medicine Research Foundation 2000.

Terwee CB, Mokkink LB, Knol DL, Ostelo RW, Bouter LM, de Vet HCW. Rating the methodological quality in systematic reviews of studies on measurement properties: a scoring system for the COSMIN checklist. Qual Life Res 2012; 21:651–7.

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.

Terwee CB, Bot SDM, de Boer MR, van der Windt DAWM, Knol DL, Dekker J, et al. Quality criteria were proposed for measurement properties of health status questionnaires. J Clin Epidemiol 2007; 60:34–42.

The PRISMA group. Preferred Reporting Items for Systematic reviews and Meta-Analyses: The PRISMA statement. 2009. http://www.prisma-statement.org/.

Toulassidharane B. Construct validity of the single arm military press (SAMP) test in a female none-patient populations. Sheffield Hallam University 2010.

Tyson S, Watson A, Moss S, Troop H, Dean-Lofthouse G, Jorritsma S, Shnnon M, (GMOM). Development of a framework for the evidence-based choice of outcome measures in neurological physiotherapy. Disability and Rehabilitation 2008;30(2):142-149.

van Tulder M, Furlan A, Bombardier C, Bouter L. Editorial Board of the Cochrane Collaboration Back Review Group. Updated method guidelines for systematic reviews in the Cochrane Collaboration Back Review Group. Spine 2003; 28:1290–9.

Walker-Bone K, Palmer K, Reading I, Coggon D, Cooper C. Prevalence and impact of musculoskeletal disorders in the upper limb in the general population. Arthritis Rheum 2004; 51:642–51.

Walker-Bone K, Byng P, Linaker C, Reading I, Coggon D, Palmer KT, et al. Reliability of the Southampton examination schedule for the diagnosis of upper limb disorders in the general population. Ann Rheum Dis 2002; 61:1103–6.

World Health Organization (WHO). International Classification of Functioning, Disability and Health: ICF. 2001. http://www.who.int/.