

RESEARCH ARTICLE

Development and Reliability of a Clinician-rated Instrument to Evaluate Function in Individuals with Shoulder Pain: A Preliminary Study

Yngve Roe^{1*}, Benjamin Haldorsen², Ida Svege³ & Astrid Bergland¹¹Faculty of Health Sciences, Oslo and Akershus University College of Applied Sciences, Oslo, Norway²Department of Physiotherapy, Martina Hansen's Hospital, Bærum, Norway³Norwegian research center for Active Rehabilitation (NAR), Oslo, Norway

Abstract

Background and Purpose. Subacromial impingement syndrome (SIS) is a common and disabling condition in the population. Interventions are often evaluated with patient-rated outcome measures. The purpose of this study was to develop a simple clinician-rated measure to detect difficulties in the execution of movement-related tasks among patients with subacromial impingement syndrome. **Method.** The steps in the scale development included a review of the clinical literature of shoulder pain to identify condition-specific questionnaires, pilot testing, clinical testing and scale construction. Twenty-one eligible items from thirteen questionnaires were extracted and included in a pilot test. All items were scored on a five-point ordinal scale ranging from 1 (no difficulty) to 5 (cannot perform). Fourteen items were excluded after pilot testing because of difficulties in standardization or other practical considerations. The remaining seven items were included in a clinical test-retest study with outpatients at a hospital. Of these, four were excluded because of psychometric reasons. From the remaining three items, a measure named Shoulder Activity Scale (summed score ranging from 3 to 15) was developed. **Results.** A total of 33 men and 30 women were included in the clinical study; age range 27–80 years. The intraclass correlation coefficient results for inter-rater reliability and test-retest reliability were 0.80 (95% CI = 0.51–0.90) and 0.74 (95% CI = 0.58–0.84), respectively. The standard error of measurement and minimal detectable change were 1.19 and 3.32, respectively. The scale was linked to the International Classification of Functioning, Disability and Health second level categories lifting and carrying objects (d430), dressing (d540), hand and arm use (d445) and control of voluntary movement (b760). **Conclusion.** The Shoulder Activity Scale showed acceptable reliability in a sample of outpatients at a hospital, rated by clinicians experienced in shoulder rehabilitation. The validity of the scale should be investigated in future studies before application to common practice. © 2013 The Authors. Physiotherapy Research International published by John Wiley & Sons Ltd.

Received 21 October 2011; Revised 5 September 2012; Accepted 5 April 2013

*Correspondence

Yngve Roe, MSc, Faculty of Health Sciences, Oslo and Akershus University College of Applied Sciences, Postboks 4 St. Olavs plass, 0130 Oslo, Norway.

E-mail: yngve.roe@hf.hio.noPublished online in Wiley Online Library (wileyonlinelibrary.com) DOI: 10.1002/pri.1555

Introduction

Shoulder pain is an umbrella term for conditions with different aetiologies and courses, and prevalence estimates have varied between 7% and 26% (Luime *et al.*, 2004, van der Heijden, 1999). Subacromial impingement syndrome (SIS) is probably the most common shoulder diagnosis, and the condition is associated with substantial loss of function (Neumann, 2010, Silva *et al.*, 2008, Lewis *et al.*, 2005, van der Windt *et al.*, 1995). SIS is describing a dysfunctional mechanism, and the alterations in movement patterns associated with the condition have been extensively analysed (Bigliani and Levine, 1997, Michener *et al.*, 2003, Neumann, 2010, Lin *et al.*, 2006, Ludewig and Cook, 2000, Lukasiewicz *et al.*, 1999). It is essential that the alterations in movement patterns are also included in functional assessments in the clinic, but few such standardized measures are available.

Reliable and valid standardized measures are important for clinical decision making and research. Patient-rated outcome measures have been recommended to evaluate interventions in patients with shoulder pain, and a number of condition-specific measures are now available (Bot *et al.*, 2004, Michener, 2011). Clinician-rated methods are also considered important in assessments, but the most commonly used measures are either a standardization of the clinical examination or physical examination tests (Constant and Murley, 1987, Richards *et al.*, 1994, Hegedus *et al.*, 2008). Although the patient-rated and clinician-rated condition-specific measures probably capture different aspects of functioning, few efforts have been made to analyse the content.

The International Classification Of Functioning, Disability and Health (ICF), provides a framework for describing and classifying the content of all measures of function (WHO, 2001). The ICF is based on an integrative model covering functioning within its components of *body functions (b)*, *body structures (s)*, *activities and participation (d)* and the *environmental (e)* and *personal factors* (not classified). The ICF classification provides categories of functioning and environmental factors that are arranged in a hierarchical fashion by using an alphanumeric coding system; the first letter referring to the component, followed by a numeric code that starts with the chapter number (e.g. *mobility*, d4-chapter), followed by the second level (e.g. *hand and arm use*, d445), third level (e.g. *reaching*, d4452) and fourth level when appropriate. Because of a generic structure, the categories at a lower level are included in the higher level categories and chapters. Procedures have been established to classify the content

of functional measures by ICF categories, regardless of their purpose, extent and by whom they are rated (Cieza *et al.*, 2002, Cieza *et al.*, 2005).

According to the ICF, the traditional clinician-rated measures may be referred to as belonging to the body functions and structures components, whereas the available patient-rated questionnaires to the activities and participation (Michener, 2011). To our knowledge, no clinician-rated measure containing content relating to the activities & participation component of the ICF has been developed. The clinician-rated measures have the advantage of directly measuring the unit of interest; they reflect the current situation and are less vulnerable to the patient's recall, language and problems with vision or literacy (Gotay, 1996). Patient and clinician ratings probably reflect different constructs, and a low to moderate correlation has been reported (Reneman *et al.*, 2002, Mannerkorpi *et al.*, 2006, Stratford and Kennedy, 2006). The aim of this study was to develop a reliable clinician-rated functional scale to measure change over time, according to the ICF component activities and participation, in patients with SIS.

Methods

Scale development

The steps in the scale development included a review of the scientific literature of shoulder pain, pilot testing, clinical testing and scale construction (Clark and Watson, 1995, Loevinger, 1957, Streiner and Norman, 2008) (Figure 1). Thirteen frequently used condition-specific questionnaires of shoulder function were identified after a review of the scientific literature. From these, 21 single items were extracted and considered eligible for pilot testing after discussions between the researchers (YR, BH and IS). All items described the execution of tasks with dynamic movements of the arm at or above shoulder level. With the participation of outpatients with shoulder pain at a hospital, the eligible items were further investigated in a pilot test. The researchers (BH and IS) and other experienced physiotherapists at the hospital participated as observers. As a result of the pilot test, 14 items that were difficult to standardize or gave little information about the patient's movement patterns were excluded. Decisions were based on agreement between all observers. In cases of disagreement, a senior member of the research group (AB) was consulted. There were no examples of such disagreement. The remaining 7 items were included in a full-scale clinical study for investigation of reliability and representation in the ICF classification.

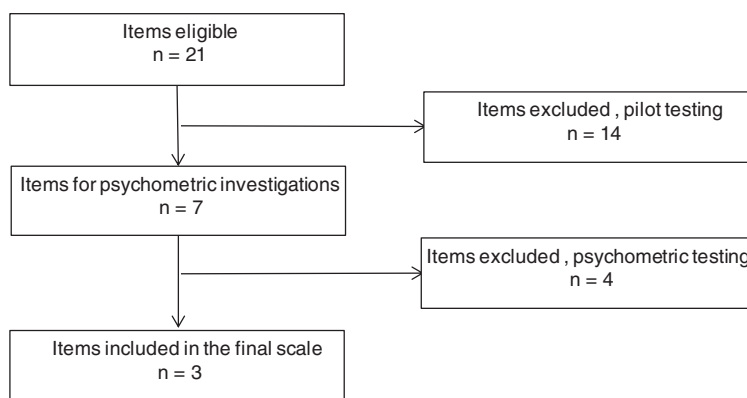


Figure 1. Flowchart of the item reduction process

To rate the magnitude of a functional problem, a five-point ordinal scale similar to the qualifiers in the ICF classification was used (WHO, 2001). The anchor points of the scale were no difficulty (1), mild difficulty (2), moderate difficulty (3), severe difficulty (4) and cannot perform (5). No definition of the term difficulty was given, as it was assumed that physical therapists experienced in shoulder rehabilitation have a common understanding of the term. The intervals between the categories were not further investigated but treated as equal in the statistical analyses.

All items were linked to second level ICF categories according to established rules (Cieza et al., 2005). Inter-item and item-to-sum correlations and representation in the ICF classification were used as exclusion criteria. A tentative summed scale named Shoulder Activity Scale (SAS) was constructed from the remaining three items and further statistically examined (Appendix 1). The included items were *lifting an object to a shelf*, *putting on a jacket* and *moving an arm sideways*. All items were weighted equal, and the scale had a possible range of 3 (no difficulties) to 15 (cannot perform). The scale was easy to administer and was in most cases completed within 5 minutes. No adverse effects from performing the SAS items were reported by the subjects or identified by the raters.

The items were linked to the ICF second level categories *lifting and carrying objects* (d430), *dressing* (d540) and *hand and arm use* (d445), respectively. The aim of the scale, to measure difficulty in terms of altered movement patterns, was linked to the *control of voluntary movement* (b760) category.

Subjects

A clinical test-retest study with outpatients attending the orthopaedic division at a hospital between December

2007 and October 2010 was conducted. The eligible patients were non-native English speakers. Inclusion criteria were primary diagnosis of SIS according to standardized criteria (Juel et al., 2008, Walker-Bone et al., 2003). Exclusion criteria were systematic inflammatory disease or generalized pain, cardiac disease, symptoms of cervical spine disease or surgery in the affected shoulder within the last 6 months.

Power analysis

A method for sample size based on the intraclass correlation coefficient (ICC), was chosen (Walter et al., 1998). The minimally acceptable ICC value ($\rho_1 = 0.7$) versus an alternative ICC value reflecting the expectations ($\rho_1 = 0.8$) was chosen. With a power of 80% ($\beta = 0.2$) and a significance level of 5%, a sample size of at least 40 patients was required (Walter et al., 1998).

Procedure and measures

Descriptive information was collected for all participants. The items were tested twice for each participant without any treatment in between. The instruction to the patients was as far as possible provided in a standardized manner and is shown in Appendix 1. The average time between baseline test and retest was 7.5 days (range 7–21). The participants were asked on the day of retest whether a substantial change in their shoulder condition had occurred since the baseline test. Participants were included in the further analyses regardless of whether a substantial change in their condition had occurred. Two independent clinicians took part in the testing at baseline, where one participated at retest. A total of five clinicians participated in the test sessions; all experienced in shoulder rehabilitation at the hospital. All clinicians

had participated in a standardized training session before conducting the test sessions.

Participants also completed the Shoulder Pain and Disability Index (SPADI) at baseline test (Roach *et al.*, 1991). SPADI is a patient-rated measure for patients with shoulder pain consisting of 13 questions, divided in the domains pain (5 items) and disability (8 items). Each item is rated on a numerical scale from 0 (best) to 10 (worst) and summed up to a domain score. Each domain score is equally weighted then added for a total percentage score ranging from 0 to 100.

Statistical analysis

The statistical analysis was conducted with the IBM SPSS Statistics 19 for windows (IBM Corporation, New York, USA) and the STATA/IC 11.1 for Mac (StataCorp LP, Lakeway Drive, Texas, USA).

The mean values or frequencies with the standard deviations (SD) were reported for the numerical or categorical variables. The association between the SAS scores and age and duration of symptoms was investigated with estimations with Pearson's product-moment correlation coefficient (r) and visual inspection of bivariate data for non-linear relations.

For further investigation of reliability, the following underlying measurement properties were chosen (Mokkink *et al.*, 2010, Terwee *et al.*, 2007): internal consistency, reliability and measurement error. Internal consistency was estimated with Cronbach's alpha. An alpha between 0.7 and 0.9 was considered fair. Consistency and unidimensionality was further investigated with inter-item correlations estimated with Pearson's product-moment correlation coefficient (Cortina, 1993). Inter-item correlations in the range of 0.15–0.50 and mean inter-item correlations of 0.40–0.50 were considered acceptable (Clark and Watson, 1995). Inter-rater reliability and test-retest reliability was estimated with the ICC. To be able to generalize the results to a population of other clinicians and because the difficulty of the items was considered to be a systematic source of variance, a two-way random effect model single measure reliability had to be chosen (Shrout and Fleiss, 1979, McGraw and Wong, 1996).

The measurement error was defined as the systematic and random error of a patient's score that was not attributed to true changes in the construct to be measured (Mokkink *et al.*, 2010). The standard error of measurement (SEM), which reflects the standard deviation of the distribution of the patient's score, with

no change in health status and no learning effect taking place, was used (Wyrwich, 2004, Weir, 2005). There are two types of SEM: $SEM_{\text{agreement}}$ and $SEM_{\text{consistency}}$. To take the systematic difference into account, the $SEM_{\text{agreement}}$ was chosen, estimated with the formula $= \sigma_x \sqrt{1 - r_{tt}}$, where (σ_x) was the pooled standard deviation of test and retest scores, and (r_{tt}) was the reliability coefficient. From the SEM value, it is possible to estimate the minimal detectable change (MDC), which is the smallest change that can be defined by the instrument beyond measurement error (de Vet *et al.*, 2006, Beckerman *et al.*, 2001). The following formula was used: $MDC = 1.96 \times \sqrt{2} \times SEM$, where 2 relates to test and retest, and 1.96 relates to the 95% confidence interval. A plot with the difference of the baseline and retest versus the mean of the sum scores was drawn (Bland and Altman, 1999). The limits of agreement (LOA) were plotted as the standard deviation of the mean difference (SD) multiplied by ± 1.96 .

All the participants signed a written consent, and the study was approved by the Norwegian Regional Committee for Ethics and conducted according to the Helsinki Declarations.

Results

Sixty-three patients, thirty women and thirty-three men participated in the clinical study. Ninety-four met the inclusion criteria, twenty-nine did not accept participation, two were excluded because of generalized pain and three dropped out between baseline test and retest. No descriptive data were recorded on eligible patients who did not accept participation. The mean age of the participants was 53.3 years (SD = 12.9). The mean duration of symptoms was 46.6 months (SD = 72.3). Thirty-eight of the participants were working, eight were sick listed and seventeen were retired, receiving disability benefit or unemployed. There were 30 cases of pain in the right shoulder, 19 in the left shoulder and 14 cases of bilateral pain. The dominant arm was affected in 30 of the cases. Five patients reported a substantial change of the condition during the test period. The mean SPADI score at baseline was 36.2 (SD = 16.6).

The item-to-item correlations ranged between 0.30 and 0.49, and the item-to-total between 0.70 and 0.82 (Table 1). The Cronbach's alpha of consistency for the SAS sum score was estimated at $\alpha = 0.86$. There were no significant correlations or non-linear associations between the participants' ages or permanence of symptoms and the SAS score.

Table 1. Significant inter-item and item-to-sum correlations with Pearson's *r* in the baseline test scores (*n* = 63)

Item	Putting on a jacket	Moving an arm sideways	Shoulder Activity Scale sum score
Lifting an object to a shelf	0.30	0.49	0.77
Putting on a jacket		0.34	0.70
Moving an arm sideways			0.82

The distribution of the scale were positively skewed as two participants had an SAS score of 3 and none above 12 (Figure 2).

The moving the arm sideways had a higher mean score than the other items, indicating that it was a more difficult task (Table 2).

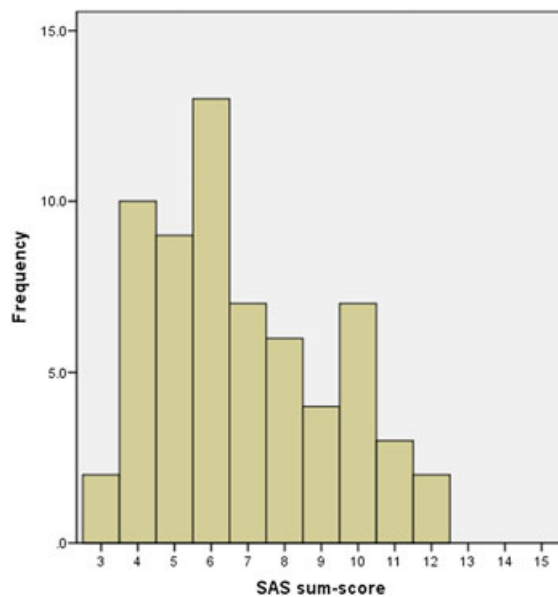
The difference between SAS test and retest was plotted against the average, with the 95% limits of agreement at -2.72 and 3.79 (Figure 3). The mean difference was 0.53 . Three out of sixty values were outside the LOA.

Discussion

The aim of this study was to develop a reliable clinician-rated functional scale to measure change over time, according to the ICF component activities and participation, in patients with SIS.

The main results of the clinical study were the findings of an inter-rater reliability and test-retest reliability of the SAS of 0.80 and 0.74 , respectively (Table 2), in line with what was expected in the power analysis. There is no commonly agreed limit for what should be considered an acceptable ICC value, but an ICC above 0.70 with the lower limit of the confidence interval above 0.60 has been proposed in clinician-rated methods (Terwee *et al.*, 2006). Even though both reliability estimates exceeded the minimum recommendations, the lower limits of the 95% confidence interval for both estimates were slightly below 0.60 . The acceptable reliability found in the current study were in line with previous findings of Westerberg and colleagues who concluded that three active motor tests had good reliability when used as functional tests in painful shoulders (Westerberg *et al.*, 1996).

The inter-item correlations (Table 1) in the final scale was within what was considered acceptable, ranging from 0.30 to 0.49 (Clark and Watson, 1995). An internal

**Figure 2.** Histogram with the distribution of Shoulder Activity Scale sum scores at the baseline test (*n* = 63)

consistency of 0.88 indicates that no items were redundant or measured other constructs. Other possible combinations of items resulted in lower alpha values. The three items were most likely not equally difficult as the item moving an arm sideways had a higher mean score (Table 2). However, the item had an acceptable inter-item correlation and item-to-total correlation (Table 1). The problems of different item-difficulty in scales are shared with other scales developed through statistical analysis based on classical test theory.

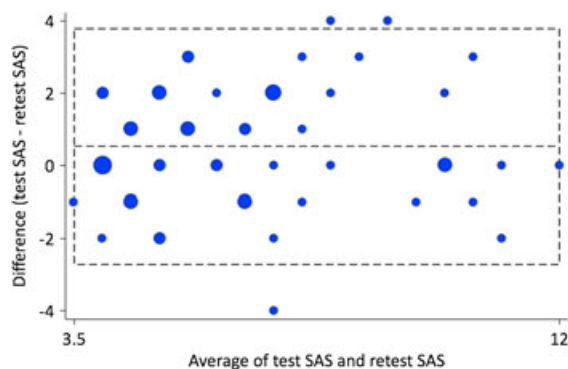
The MDC for the SAS was estimated to 3.30 (Table 2). The interpretation is that individual changes in the sum score of 1–3 points can be due to systematic or random errors. In classical test theory, the MDC is considered a stable property of the instrument, and a change in the sum score of 4 or higher should thus be considered real but not necessarily clinically relevant (de Vet *et al.*, 2006). The MDC should not be interchanged with the minimal important difference, which refers to the benefit of treatment in a specific population (de Vet *et al.*, 2006, de Vet and Terwee, 2010). Controversy exists whether the benefit of treatment estimates should be derived from distribution-based or anchor-based methods. Norman and colleagues found consistent evidence that the minimal important difference equals close to half of an SD at baseline in a systematic literature review where both anchor-based and distribution-based methods had been used (Norman *et al.*, 2003). Furthermore, Wyrwich suggested

Table 2. Reliability estimates ($n=60$) with pooled test-retest mean, range and inter-rater reliability, test-retest reliability, standard error of measurement (SEM), minimal detectable change (MDC) and effect size for single items (1–5) and sum score (3–15)

Item	Mean (SD)	Range	ICC inter-rater (95% CI)	ICC test-retest (95% CI)	SEM	MDC
Lifting an object to a shelf	1.87 (0.98)	1–5	0.66 (0.35–0.82)	0.59 (0.40–0.73)	0.61	1.69
Putting on a jacket	1.94 (0.98)	1–5	0.71 (0.42–0.85)	0.55 (0.35–0.71)	0.62	1.72
Moving an arm sideways	3.00 (1.15)	1–5	0.75 (0.61–0.84)	0.84 (0.75–0.90)	0.45	1.25
SAS sum score	6.81 (2.38)	13–12	0.80 (0.51–0.90)	0.74 (0.58–0.84)	1.19	3.30

a one-to-one relation between the minimal important difference and the SEM (Wyrwich, 2004). Estimates based on the aforementioned distribution-based methods resulted in a minimal important difference of 1.19 in both cases. According to the estimation methods recommended by Norman and Wyrwich, an SAS sum score of at least 4 is also clinically important.

The participants had a high functional level measured with SPADI, compared with other studies including patients with subacromial conditions (Ekeberg *et al.*, 2008, Williams *et al.*, 1995). There were only two patients with the lowest SAS score of 3, and none with the sum scores 13–15 (Figure 2). Even though the distribution was obviously skewed, this is less than the 15% normally considered a floor effect (Terwee *et al.*, 2007). A skewed distribution however should not necessarily be considered a problem in functional scales but rather a common and logical manifestation of the underlying construct (Streiner and Norman, 2008). The LOA-plot (Figure 3) gives a graphical expression of the ability of an instrument to replicate observations, and the differences should ideally be close to zero (Bland and Altman, 1999). The plot gives a visual indication of a slightly higher retest score among most participants, consistent for both low and high SAS average scores.

**Figure 3.** Intra-individual differences ($n=60$) plotted against the difference between test and retest scores on Shoulder Activity Scale. The central horizontal line represents the mean difference, whereas the flanking lines represent the 95% limits of agreement

The items in SAS were linked to ICF categories from the *mobility* (d4-chapter) or *self-care* (d5-chapter) of the *activities and participation* component, and the aim of the scale was linked to the *neuromusculoskeletal and movement-related functions* (b7-chapter) of the *body functions* component (WHO, 2001). To our knowledge, no other similar clinician-rated activity scale exists. The standardized clinical examination methods and the physical examination tests commonly used in the assessments have no content relating to the *activities and participation* component of the ICF (Constant and Murley, 1987, Hegedus *et al.*, 2008, Richards *et al.*, 1994). The FiT-HaNSA-test focuses on muscle endurance, which is also covered by the *body functions* component (MacDermid *et al.*, 2007). Hence, the test probably measures a different construct than the SAS.

The SAS needs to be validated before implemented into clinic. Nevertheless, the current study may contribute to increase the attention on the content of functional assessments in patients with shoulder pain. The study may facilitate a further use of the ICF to classify functional measures. Future work should further investigate how standardized clinician-rated measures may be implemented in functional assessments and how they relate to the patient-rated measures.

Study limitations

First, the SAS is based on the assumption that clinicians have a common understanding of the term *difficulty*. Although the assumption is supported by the findings of the current study, it may have contributed that all the raters were working at the same hospital. No commonly agreed on guidelines for assessments of shoulder pain yet exists. Second, the treatment of ordinal data as numerical in the statistical analyses may be questioned, because no investigations of the intervals between the anchor points had been conducted. The approach was chosen because of the fact that most statistical methods used in psychometric evaluations require numerical

data (Streiner and Norman, 2008). Third, it should be recognized that the test was applied to a non-native English-speaking population, and it is thus possible that native English-speaking patients might interpret the instructions differently.

Conclusions

The SAS seems to be a reliable clinician-rated instrument to measure functional change in patients with SIS. A change score of at least 4 points is required for evaluation of individual patients. Time of administration was less than 5 minutes, and no specialized equipment is required. The content of the scale is covered by the mobility (d4-chapter) and self-care (d5-chapter) of the ICF. The validity of the scale needs to be established before it is applied to common practice.

Acknowledgements

We thank Gerty Lund and Ingrid Walter who contributed in the data collection, and all the patients who participated in the study.

The study was supported by The Norwegian Fund for Postgraduate Training in physiotherapy and Bergesens Almennyttige Stiftelse (Bergesen's Foundation).

We certify that no party having a direct interest in the results of the research supporting this article has or will confer a benefit on us or on any organization with which we are associated and all financial and material support for this research and work are clearly identified in the title page of the manuscript.

REFERENCES

- Beckerman H, Roebroek ME, Lankhorst GJ, Becher JG, Bezemer PD, Verbeek AL. Smallest real difference, a link between reproducibility and responsiveness. *Quality of Life Research* 2001; 10: 571–578.
- Bigliani LU, Levine WN. Subacromial impingement syndrome. *The Journal of Bone and Joint Surgery. American Volume* 1997; 79: 1854–1868.
- Bland JM, Altman DG. Measuring agreement in method comparison studies. *Statistical Methods in Medical Research* 1999; 8: 135–160.
- Bot SD, Terwee CB, van der Windt DA, Bouter LM, Dekker J, de Vet HC. Clinimetric evaluation of shoulder disability questionnaires: a systematic review of the literature. *Annals of the Rheumatic Diseases* 2004; 63: 335–341.
- Cieza A, Brockow T, Ewert T, Amman E, Kollerits B, Chatterji S, Ustun TB, Stucki G. Linking health-status measurements to the international classification of functioning, disability and health. *Journal of Rehabilitation Medicine* 2002; 34: 205–210.
- Cieza A, Geyh S, Chatterji S, Kostanjsek N, Ustun B, Stucki G. ICF linking rules: an update based on lessons learned. *Journal of Rehabilitation Medicine* 2005; 37: 212–218.
- Clark LA, Watson D. Constructing validity: basic issues in objective scale development. *Psychological Assessment* 1995; 7: 309–319.
- Constant CR, Murley AH. A clinical method of functional assessment of the shoulder. *Clinical Orthopaedics and Related Research* 1987; (214): 160–164.
- Cortina JM. What is coefficient alpha? An examination of theory and applications. *Journal of Applied Psychology* 1993; 78: 98–104.
- de Vet HC, Terwee CB, Ostelo RW, Beckerman H, Knol DL, Bouter LM. Minimal changes in health status questionnaires: distinction between minimally detectable change and minimally important change. *Health and Quality of Life Outcomes* 2006; 4: 54.
- de Vet HCW, Terwee CB. The minimal detectable change should not replace the minimal important difference. *Journal Of Clinical Epidemiology* 2010; 63: 804–805; author reply 806.
- Ekeberg OM, Bautz-Holter E, Tveita EK, Keller A, Juel NG, Brox JI. Agreement, reliability and validity in 3 shoulder questionnaires in patients with rotator cuff disease. *BMC Musculoskeletal Disorders* 2008; 9: 68–76.
- Gotay CC. Patient-reported assessments versus performance-based tests. In: Spilker B (ed.), *Quality of Life and Pharmacoeconomics in Clinical Trials*. Philadelphia: Lippincott–Raven Publishers, 1996.
- Hegedus EJ, Goode A, Campbell S, Morin A, Tamaddoni M, Moorman CT, III, Cook C. Physical examination tests of the shoulder: a systematic review with meta-analysis of individual tests. *British Journal of Sports Medicine* 2008; 42: 80–92.
- Juel NG, Brox JI, Thingnaes K, Bjornerheim R, Bryde P, Villerso K, Aakhus S. Musculoskeletal pain in ultrasound operators. *Tidsskrift for den Norske laegeforening: tidsskrift for praktisk medicin, ny raekke* 2008; 128: 2701–2705.
- Lewis JS, Green A, Wright C. Subacromial impingement syndrome: the role of posture and muscle imbalance. *Journal of Shoulder and Elbow Surgery* 2005; 14: 385–392.
- Lin JJ, Lim HK, Soto-Quijano DA, Hanten WP, Olson SL, Roddey TS, Sherwood AM. Altered patterns of muscle activation during performance of four functional tasks in patients with shoulder disorders: interpretation from

- voluntary response index. *Journal Of Electromyography And Kinesiology: Official Journal Of The International Society Of Electrophysiological Kinesiology* 2006; 16: 458–468.
- Loevinger J. Objective tests as instruments of psychological theory. *Psychological Reports* 1957; 3: 635–694.
- Ludewig PM, Cook TM. Alterations in shoulder kinematics and associated muscle activity in people with symptoms of shoulder impingement. *Physical Therapy* 2000; 80: 276–291.
- Luime JJ, Koes BW, Hendriksen IJ, Burdorf A, Verhagen AP, Miedema HS, Verhaar JA. Prevalence and incidence of shoulder pain in the general population; a systematic review. *Scandinavian Journal of Rheumatology* 2004; 33: 73–81.
- Lukasiewicz AC, McClure P, Michener L, Pratt N, Sennett B. Comparison of 3-dimensional scapular position and orientation between subjects with and without shoulder impingement. *The Journal of Orthopaedic and Sports Physical Therapy* 1999; 29: 574–583; discussion 584–6.
- MacDermid JC, Ghobrial M, Quirion KB, St-Amour M, Tsui T, Humphreys D, McCluskie J, Shewayhat E, Galea V. Validation of a new test that assesses functional performance of the upper extremity and neck (FIT-HaNSA) in patients with shoulder pathology. *BMC Musculoskeletal Disorders* 2007; 8: 42.
- Mannerkorpi K, Svantesson U, Broberg C. Relationships between performance-based tests and patients' ratings of activity limitations, self-efficacy, and pain in fibromyalgia. *Archives Of Physical Medicine And Rehabilitation* 2006; 87: 259–264.
- McGraw KO, Wong S. Forming inferences about some intraclass correlation coefficients. *Psychological Methods* 1996; 1: 30–46.
- Michener LA. Patient- and clinician-rated outcome measures for clinical decision making in rehabilitation. *Journal of Sport Rehabilitation* 2011; 20: 37–45.
- Michener LA, McClure PW, Karduna AR. Anatomical and biomechanical mechanisms of subacromial impingement syndrome. *Clinical biomechanics* 2003; 18: 369–379.
- Mokkink LB, Terwee CB, Patrick DL, Alonso J, Stratford PW, Knol DL, Bouter LM, de Vet HCW. The COSMIN study reached international consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes. *Journal Of Clinical Epidemiology* 2010; 63: 737–745.
- Neumann DA. *Kinesiology of the Musculoskeletal System. Foundations for Rehabilitation*. Mosby Elsevier: St. Louis, Mosby, 2010.
- Norman GR, Sloan JA, Wyrwich KW. Interpretation of changes in health-related quality of life: the remarkable universality of half a standard deviation. *Medical Care* 2003; 41: 582–592.
- Reneman MF, Jorritsma W, Schellekens JM, Goeken LN. Concurrent validity of questionnaire and performance-based disability measurements in patients with chronic nonspecific low back pain. *Journal of Occupational Rehabilitation* 2002; 12: 119–129.
- Richards RR, An K-N, Bigliani LU, Friedman RJ, Gartsman GM, Gristina AG, Iannotti JP, Mow VC, Sidles JA, Zuckerman JD. A standardized method for the assessment of shoulder function. *Journal of Shoulder and Elbow Surgery* 1994; 3: 347–352.
- Roach KE, Budiman-Mak E, Songsirdej N, Lertratanakul Y. Development of a shoulder pain and disability index. *Arthritis Care And Research: The Official Journal Of The Arthritis Health Professions Association* 1991; 4: 143–149.
- Shrout PE, Fleiss JL. Intraclass correlations: uses in assessing rater reliability. *Psychological Bulletin* 1979; 86: 420–428.
- Silva L, Andreu JL, Munoz P, Pastrana M, Millan I, Sanz J, Barbadillo C, Fernandez-Castro M. Accuracy of physical examination in subacromial impingement syndrome. *Rheumatology* 2008; 47: 679–683.
- Stratford PW, Kennedy DM. Performance measures were necessary to obtain a complete picture of osteoarthritic patients. *Journal of Clinical Epidemiology* 2006; 59: 160–167.
- Streiner DL, Norman GR. *Health Measurement Scales. A Practical Guide to Their Development and Use*. Oxford: Oxford University Press, 2008.
- Terwee CB, Bot SD, de Boer MR, van der Windt DA, Knol DL, Dekker J, Bouter LM, de Vet HC. Quality criteria were proposed for measurement properties of health status questionnaires. *Journal of Clinical Epidemiology* 2007; 60: 34–42.
- Terwee CB, Mokkink LB, Steultjens MP, Dekker J. Performance-based methods for measuring the physical function of patients with osteoarthritis of the hip or knee: a systematic review of measurement properties. *Rheumatology (Oxford, England)* 2006; 45: 890–902.
- van der Heijden GJ. Shoulder disorders: a state-of-the-art review. *Best Practice & Research. Clinical Rheumatology* 1999; 13: 287–309.
- van der Windt DA, Koes BW, de Jong BA, Bouter LM. Shoulder disorders in general practice: incidence, patient characteristics, and management. *Annals Of The Rheumatic Diseases* 1995; 54: 959–964.
- Walker-Bone KE, Palmer KT, Reading I, Cooper C. Criteria for assessing pain and nonarticular soft-tissue rheumatic disorders of the neck and upper limb. *Seminars in Arthritis and Rheumatism* 2003; 33: 168–184.
- Walter SD, Eliasziw M, Donner A. Sample size and optimal designs for reliability studies. *Statistics in Medicine* 1998; 17: 101–110.

Weir JP. Quantifying test-retest reliability using the intraclass correlation coefficient and the SEM. *Journal of Strength and Conditioning Research* 2005; 19: 231–240.

Westerberg CE, Solem-Bertoft E, Lundh I. The reliability of three active motor tests used in painful shoulder disorders. Presentation of a method of general applicability for the analysis of reliability in the presence of pain. *Scandinavian Journal Of Rehabilitation Medicine* 1996; 28: 63–70.

WHO. *International Classification of Functioning, Disability and Health: ICF*. Geneva: World Health Organization, 2001.

Williams JW, Jr., Holleman DR, Jr., Simel DL. Measuring shoulder function with the Shoulder Pain and Disability Index. *Journal of Rheumatology* 1995; 22: 727–732.

Wyrwich KW. Minimal important difference thresholds and the standard error of measurement: is there a connection? *Journal of Biopharmaceutical Statistics* 2004; 14: 97–110.

Appendix 1. Shoulder Activity Scale

Test	Procedure	Instruction	Score (circle the most relevant)				
			No difficulty	Mild difficulty	Moderate difficulty	Severe difficulty	Cannot perform
1. Lifting an object to a shelf	From a standing or sitting position, the subject lifts a 1-kg object from a table to a high shelf. The task is repeated three times without a break. The height of the shelf should be slightly above the subject's head, and the difference in height between the table and the shelf is at least 0.7 m.	<i>Lift the object up on the shelf and back on the table three times.</i>	1	2	3	4	5
2. Putting on a jacket	From a standing or sitting position, the subject puts on a jacket with the healthy arm in the first sleeve and then off beginning with the painful arm. The jacket should be medium tight and made of non-stretchy material.	<i>Put on the jacket with the healthy arm in the first sleeve and take it off with the painful arm first.</i>	1	2	3	4	5
3. Moving an arm sideways	From a sitting position, with approximately 90° angle in the hip and knee, the subject lifts a 2-kg object with a straight and approximately 90° internal rotated arm, from a table in front and to the height of the shoulder. The arm is now at 90° flexion, internal rotated in a sagittal plane. The straight arm is abducted to the frontal plane, and adducted to the sagittal plane without allowing any variation in the height or the rotation of the arm. The task is repeated once without a break.	<i>Lift the object up from the desk to shoulder height with a straight arm. Keep the upper body stable. Move the object sideways until the arm is outside the shoulder, and then back to forward position. Keep the arm at shoulder level and straight through the movement. The task is repeated once without a break.</i>	1	2	3	4	5
Sum-score	1 + 2 + 3 = ____ points						