

RESEARCH ARTICLE

Open Access



# Responsiveness and minimal important change of the Norwegian version of the Disabilities of the Arm, Shoulder and Hand questionnaire (DASH) in patients with subacromial pain syndrome

Tarjei Rysstad<sup>1\*</sup> , Yngve Røe<sup>1</sup>, Benjamin Haldorsen<sup>2</sup>, Ida Svege<sup>3</sup> and Liv Inger Strand<sup>4</sup>

## Abstract

**Background:** The Disabilities of the Arm, Shoulder, and Hand questionnaire (DASH) is a valid and reliable patient-reported outcome measure (PROM). It was designed to measure physical disability and symptoms in patients with musculoskeletal disorders of the upper extremity, and is one of the most commonly used PROMs for patients with shoulder pain. The aim of this study was to examine responsiveness, the smallest detectable change (SDC) and the minimal important change (MIC) of the DASH, in line with international (COSMIN) recommendations.

**Methods:** The study sample consisted of 50 patients with subacromial pain syndrome, undergoing physical therapy for 3–4 months. Responsiveness to change was examined by calculating area under the receiver operating characteristic curves (AUC) and testing a priori-formulated hypothesis regarding correlations with changes in other instruments that measuring the same construct. The SDC was calculated using a test re-test protocol, and the MIC was calculated by the anchor-based MIC distribution. MIC values for patients with low and high baseline scores were also calculated.

**Results:** DASH appeared to be responsive, as it was able to distinguish patients who reported to be improved from those unchanged (AUC 0.77). All of the hypotheses were accepted. The SDC was 11.8, and the MIC was 4.4.

**Conclusion:** This study shows that the Norwegian version of the DASH has good responsiveness to change and may thus be recommended to measure outcome in patients with shoulder pain in Norway.

**Keywords:** Dash, Responsiveness, Minimal important change, MIC, Cosmin, Physical therapy

## Background

Shoulder pain is common in the general population with a reported 1-year prevalence of 5–47% [1]. Subacromial pain syndrome is the most common diagnosis for patients with shoulder pain [2]. Numerous patient-reported outcome measures (PROMs) are available for the evaluation of shoulder disorders [3]. PROMs have become standard instruments in intervention studies, and are increasingly being used in clinical practice to

capture patients' self-reported function, disability and health [4]. The use of PROMs in clinical practice can lead to improved patient-clinician communication and enhance patient care and outcomes [5]. It is essential that PROMs demonstrate acceptable psychometric properties [6]. To obtain consensus of how measurement properties are defined and tested, an international Delphi process has been carried out to develop methodological guidelines, called the Consensus-based Standards for the development of Measurement INstruments (COSMIN) [7, 8]. A prerequisite for a PROM to be used for evaluation in both clinical practice and research, is the instrument's responsiveness [9]. Responsiveness is defined as "the ability of an

\* Correspondence: rysstad@live.no

<sup>1</sup>Faculty of Health Sciences, Oslo and Akershus University College of Applied Sciences, Postboks 4, St Olavs Plass, Oslo, Norway

Full list of author information is available at the end of the article



instrument to detect change over time in the construct to be measured”, and is considered an aspect of validity [7]. To interpret change scores of a PROM two benchmarks are required: the smallest detectable change (SDC) and the minimal important change (MIC), which the COSMIN group defines respectively as “the smallest change that can be detected by the instrument, beyond measurement error” and “the smallest change in the construct to be measured which patients perceive as important” [9].

The DASH is a region specific PROM, and is according to Roy et al. [10] the most commonly used questionnaire for assessing shoulder function. It was developed to measure physical disability and symptoms of single or multiple disorders in the upper limb region [11], and has been translated into several languages including Norwegian [12]. The DASH has shown acceptable reliability, validity and responsiveness across a number of shoulder pathologies [10, 13].

However, according to the COSMIN panel inadequate methods for examining responsiveness and MIC have been applied, such as effect size (ES), standardized response mean (SRM) and distribution-based approaches [14]. The COSMIN panel claims that these methods measures the magnitude of an effect, and not the instruments ability to detect change over time or the importance of the change [7, 9]. Roy et al. [10] pointed out the need to conduct more studies regarding the estimation of MIC in shoulder disability scales. Since the responsiveness and MIC of a PROM may vary by population and context [15], it is important to examine these measurement properties in different samples of patients.

To our knowledge, responsiveness of the DASH has not previously been examined in line with the COSMIN recommendations. Hence, the purpose of this study was to assess responsiveness, and to determine and compare the SDC and MIC of the Norwegian version of the questionnaire following these guidelines.

## Methods

### Study participants and procedure

This clinimetric study included patients with subacromial pain syndrome (SPS). The patients were recruited from an outpatient clinic for physical therapy and rehabilitation at X, from December 2007 to October 2010. The patients were referred to “usual physical therapy” in primary health care, partly financed by the public health insurance service. The inclusion criteria was a diagnosis of SPS given by an orthopaedic surgeon based on clinical findings and symptoms, including anterior-lateral shoulder pain worsening during elevation of the arm and overhead activities, positive isometric abduction and positive impingement sign [16]. Exclusion criteria were systematic disease or generalized pain, cardiac disease, symptoms of cervical spine disease or surgery in the

affected shoulder within the last 6 months. Patients were also excluded if they were unable to read or speak Norwegian fluently. Ethical approval was obtained by the Norwegian Regional Committee for Ethics, and all the participants signed a written consent form.

The assessments for the present study were completed at three different time-points: T1 (baseline), T2 (1–2 weeks after baseline) and T3 (3–4 months follow-up). At T1 they completed the DASH, Shoulder Pain and Disability Index (SPADI), Short Form 36 Health Survey (SF-36), Numeric Pain Rating Scale (NPRS), and provided demographic data. Active range of motion (AROM) (shoulder abduction, flexion, and internal and external rotation) was measured with a goniometer by a physiotherapist. At T2, the patients only filled out the DASH and SPADI. After completed treatment (T3) they filled out DASH, SPADI, SF-36, and NPRS, and AROM was measured. Perceived recovery was reported by the patients on a 3-level ordinal scale. Mean DASH scores of T1 and T2 were calculated as a basis for measuring change.

### The Disabilities of the Arm, Shoulder and Hand questionnaire (DASH)

The DASH consists of a 30-item scale that contains 21 physical function items, five symptom items and four social role items. Each item has five response options concerning the patient’s symptom severity and function of the upper extremity in activity during the previous week. The scores range from 0 to 100, where 100 reflects the most severe disability. The DASH was considered incomplete if more than 10% (missing rule) of the items scores were missing. Missing items were imputed as the average of the remaining items [17]. A previous study based on the same sample of patients demonstrated acceptable test-retest reliability, internal consistency and construct validity of the Norwegian version of the DASH [18].

### Other assessment tools

*The Shoulder Pain and Disability Index (SPADI)* measures shoulder pain and disability, and contains 13 items in two domains: pain (five items) and disability (eight items). The total score ranges from 0 to 100, lower scores indicating less pain and disability [19]. According to systematic reviews, the SPADI has demonstrated high reliability [10, 20] and satisfactory validity in large groups of patients with shoulder pain [21], but its validity and reliability has been questioned in a recent review [22]. The Norwegian version of SPADI used in this study has shown adequate validity and reliability [23].

*The Short Form 36 Health Survey (SF-36)* measures health in eight domains: physical functioning (PF), role-physical (RP), bodily pain (BP), perception of general health (GH), energy and vitality (VT), social functioning

(SF), role limitation due to emotional problems (RE), and mental health (MH). The results range from the worst outcome (0) to the best outcome (100). The SF-36 was used as a generic measure of health, recommended to supplement condition-specific measures in shoulder patients [24]. The scoring was carried out following the published guidelines [25]. The Norwegian version of SF-36 has demonstrated acceptable psychometric properties [26].

*Numeric Pain Rating Scale (NPRS)* captures the patient's level of pain. It is an 11-point scale that is anchored on the left with "no pain", and on the right with "worst pain possible". The patients were asked to rate their current pain within the last 24 h. NPRS has shown good reliability and responsiveness, with reported MIC values varying from 1.1 to 2.17 in patients with shoulder pain [27, 28].

*Perceived recovery.* At the 3–4 months' follow-up (T3) the patients were asked to rate on a 3-level ordinal scale whether their condition had improved, was unchanged or had deteriorated since baseline. The use of a change scale is common, while designs and the number of scoring alternatives may vary. In a recent review change scales were shown to be reliable and responsive [29], and excellent test-retest reliability was demonstrated in patients with musculoskeletal disorders [30], although validity has been questioned [31, 32].

#### Assessment of responsiveness

Both criterion and construct approaches were used to examine responsiveness [9]. As a criterion approach, the perceived recovery after treatment was used as an anchor (gold standard) for important changes in DASH scores. A dichotomous variable (the 'improved' versus 'unchanged') from the measure of change (perceived recovery) was used to examine the discriminate ability of change scores from the questionnaires, using the receiver operating curves (ROC) method. The area under the ROC curve (AUC) was used as an indicator of responsiveness. AUC is a measure of the instrument's ability to discriminate between two groups according to an external gold standard (here: perceived recovery). For sufficient responsiveness, an AUC over 0.70 is recommended [9].

The construct approach includes a priori hypotheses of expected associations between scores of DASH and other assessment tools that measure more or less the same construct (NPRS, SPADI, SF-36 and AROM). The strength of the positive correlation coefficients were interpreted according to Munro [33]: little, if any (.00–.25), low (.26–.49), moderate (.50–0.69), high (.70–0.89) and very high correlation (.90–1.00). Negative correlations were interpreted in a similar way. We used criteria described by de Boer et al. [34], which rates responsiveness as high if

less than 25% of the hypotheses are refuted, moderate if 25–50% are refuted and poor if more than 50% are refuted. The hypotheses tested are explained in detail and presented in Table 1.

#### Measurement error by Smallest Detectable Change (SDC)

The patients did not receive any treatment between T1 and T2. These two measurement points were used to determine measurement error. Measurement error can be expressed by the standard error of measurement (SEM) and the SDC. The COSMIN-group defines SEM as the standard deviation (SD) around a single measurement [9], and it was calculated as the square root of the within-subject total variance of an ANOVA analysis [35]. The intra-class correlation coefficient (ICC) was computed using a two-way mixed effects model for total agreement. The SDC was calculated as  $1.96 \times \sqrt{2} \times \text{SEM}$ . If change is above this value in individual patients one can be 95% confident that it is not caused by measurement error [36].

#### Minimal Important Change (MIC)

To explore interpretability of change scores, the SDC was compared to the MIC. To distinguish clinically important change from measurement error, the MIC

**Table 1** A priori hypotheses to assess DASH responsiveness in patients with subacromial pain syndrome ( $n = 50$ )

No.	Hypotheses and rationales for the hypotheses	
1	The correlation between change scores of DASH and NPRS is moderate ( $r = 0.50$ – $0.69$ ). Disability and pain are different, but related constructs as shown in previous studies [54, 55].	Yes
2	The correlation between change scores of DASH and SPADI is high ( $r = 0.70$ – $0.89$ ). Both questionnaires assess disability in similar shoulder pain conditions, and high correlation coefficients have been shown in previous studies [13, 56].	Yes
3	The correlation between change scores of DASH and the subdomain SPADI function is higher than between change scores of DASH and the subdomain SPADI pain. This hypothesis is based on previous studies showing that DASH has a lower emphasis on pain than on function and disability [13, 57].	Yes
4	The correlation between change scores of DASH and the SF-36 subdomain Physical functioning (PF) is higher than between change scores of DASH and the SF-36 subdomain Bodily pain (BP). This hypothesis has the same rationale as hypothesis no. 3.	Yes
5	The correlation between change scores of DASH and AROM abduction is moderately and negatively correlated ( $r = -.50$ – $0.69$ ). This hypothesis is based on findings in previous studies [46, 58].	Yes
6	The correlation between change scores of DASH and AROM of abduction is higher than the correlation between change scores of DASH and other AROM movements. Abduction is a typical impairment in sub-acromial pain syndrome [59].	Yes
Number of accepted hypotheses (%)		6 (100)

DASH disabilities of the arm, shoulder and hand questionnaire, NPRS numeric pain rating scale, SPADI shoulder pain and disability index, SF-36 The Short Form 36 Health Survey (SF-36), AROM active range of motion

should be greater than the SDC. The MIC was determined by using visual anchor-based MIC distribution based on the ROC method [37]. The perceived recovery was used as an anchor, dichotomized as 'improved' and 'unchanged'. The subgroup 'deteriorated' was not included in the ROC curve analysis. The sensitivity and 1-specificity values from the 'improved' and 'unchanged' group were plotted on the y- and x-axis to distinguish the two groups. To define the MIC for the DASH a ROC cut-off point was detected, by finding the maximized value of both sensitivity and 1-specificity [37]. The anchor was considered acceptable if a minimum correlation of 0.5 was found between the change scores of the DASH and the anchor [38].

In recent studies the MIC of PROMs has been found to vary depending on the baseline scores [39, 40], and MIC was therefore also calculated for subgroups depending on the median of the baseline scores.

#### Data analysis

The data was assumed to be normally distributed if there was no or minimal difference between the mean and median value, confirmed by histograms, by Q-Q plot and by the Shapiro-Wilk test. Pearson and Spearman correlation coefficients were used as appropriate depending on fulfillment of normality criteria. Floor and ceiling effects were considered to be present if the lowest or highest possible score was achieved by more than 15% of the patients [6]. The analysis was performed using the software IBM SPSS version 22 for Mac.

#### Results

Ninety-four patients met the inclusion criteria, while 29 (31.0%) were unwilling or unable to participate, and two were excluded because of generalized pain. Of the 63 subjects, 13 dropped out before follow-up assessment after treatment (between baseline test and 3–4 months follow-up). The final study population consisted of 22 women and 28 men, with a mean age of 54.4 years (Table 2). The mean treatment length was 15.8 weeks. Table 3 presents the patient characteristic according to the perceived recovery in question.

At baseline, two of the DASH items (0.13%) were identified as missing. Five items were missing (0.33%) at T3. All of the completed DASH questionnaires had at least 27 of the 30 items answered (missing rule), and the missing items were imputed as the average of the remaining items.

#### Responsiveness

Figure 1 presents the ROC curves generated for the DASH. Based on the anchor (perceived recovery), a total of 30 patients reported "improvement" and 13 "unchanged" (Table 2). The AUC of the DASH was 0.77 (95% CI: 0.63, 0.92).

**Table 2** Demographic data of patients with subacromial pain syndrome

	Follow-up	Drop-outs
Number	50	13
Gender		
Female (%)	22 (44)	8 (61.5)
Male (%)	28 (56)	5 (38.5)
Age (SD)	54.4 (12.9)	48.8 (12.5)
Affected side		
Left (%)	26 (52)	4 (30.8)
Right (%)	14 (28)	5 (38.5)
Both (%)	10 (20)	4 (30.8)
Symptom duration (months)		
Mean (SD)	44.3 (72.8)	55.7 (72.5)
Working/student full time (%)	23 (46)	6 (46.2)
Sick listed 100% (%)	6 (12)	3 (23.1)
Partial sick listed (%)	5 (10)	3 (23.1)
Retired (%)	13 (26)	1 (7.7)
Receiving disability benefit (%)	2 (4)	0 (0)
Unemployed (%)	1 (2)	0 (0)

Table 4 displays correlations coefficients between change scores for the instruments. All hypotheses regarding the DASH were confirmed, and are presented in Table 1.

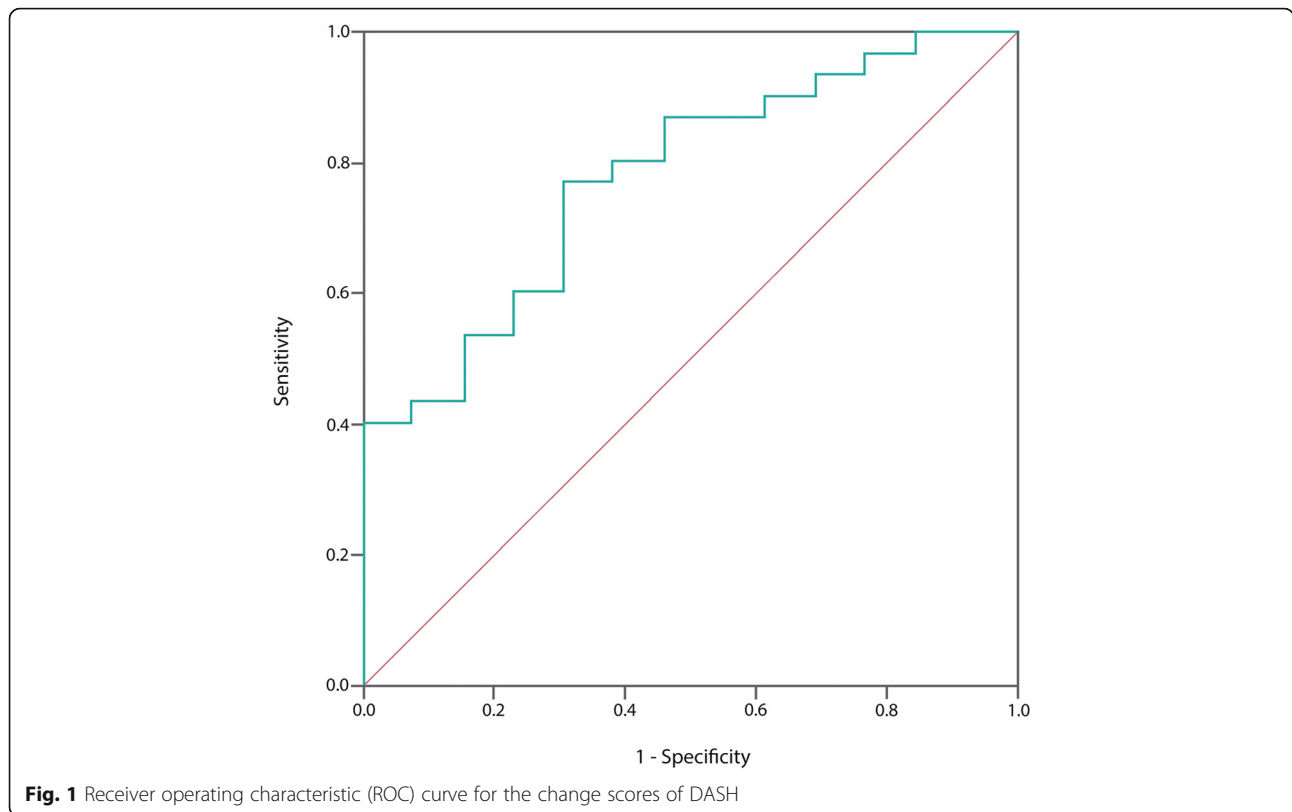
No floor or ceiling effects were observed for baseline or follow-up scores. The score distribution of the DASH is presented in Fig. 2.

**Table 3** Characteristics of the improved, unchanged and deteriorated patients at 3–4 months follow-up according to the anchor ( $n = 50$ )

	Improved <sup>a</sup>	Unchanged <sup>a</sup>	Deteriorated <sup>a</sup>
Number (%)	30 (60)	13 (26)	7 (14)
Gender (female/male) (%)	46.7/53.3	23.1/76.9	71.4/28.6
Age	53.7 ± 11.7	58.5 ± 15.0	50.1 ± 13.6
DASH Baseline	27.4 ± 12.1	29.7 ± 16.9	28.0 ± 14.3
DASH Follow-up	16.8 ± 13.3	30.6 ± 21.5	38.8 ± 13.3
DASH Change score	10.6 ± 11.7	- 0.9 ± 10.1	- 10.8 ± 8.8
SPADI Baseline	31.9 ± 11.7	40.8 ± 20.1	41.6 ± 19.0
SPADI Follow-up	18.6 ± 15.6	36.1 ± 20.2	54.5 ± 30.0
SPADI Change score	13.3 ± 14.7	4.7 ± 4.7	- 12.9 ± 16.3
NPRS Baseline	3.7 ± 1.9	3.9 ± 2.0	4.6 ± 1.5
NPRS Follow-up	2.5 ± 1.9	3.5 ± 2.3	6.6 ± 2.1
NPRS Change score	1.2 ± 2.0	0.5 ± 1.0	- 2.0 ± 2.2

DASH disabilities of the arm, shoulder and hand questionnaire, SPADI shoulder pain and disability index, NPRS numeric pain rating scale

<sup>a</sup>Data are means ± standard deviations unless denoted otherwise



**Fig. 1** Receiver operating characteristic (ROC) curve for the change scores of DASH

**Table 4** Pearson (*r*) or Spearman (*r<sup>s</sup>*) correlation between the change scores of the DASH and other measurement tools (*n* = 50)

	DASH (0–100)
Assessment tools	(0–100)
NPRS (0–10)	<i>r<sup>s</sup></i> = 0.69
SPADI (0–100)	<i>r</i> = 0.82
SPADI Function (0–100)	<i>r</i> = 0.82
SPADI Pain (0–100)	<i>r</i> = 0.77
SF-36 Physical functioning (PF) (0–100)	<i>r<sup>s</sup></i> = -0.61
SF-36 Physical limitations (RP) (0–100)	<i>r<sup>s</sup></i> = -0.60
SF-36 Bodily pain (BP) (0–100)	<i>r<sup>s</sup></i> = -0.53
SF-36 General health perception (GH) (0–100)	<i>r<sup>s</sup></i> = -0.17
SF-36 Vitality (VT) (0–100)	<i>r<sup>s</sup></i> = -0.30
SF-36 Limitation in social activities (SF) (0–100)	<i>r<sup>s</sup></i> = -0.33
SF-36 Emotional problems (RE) (0–100)	<i>r<sup>s</sup></i> = -0.46
SF-36 Mental health (MH) (0–100)	<i>r<sup>s</sup></i> = -0.44
AROM Abduction	<i>r<sup>s</sup></i> = -0.61
AROM Flexion	<i>r<sup>s</sup></i> = -0.41
AROM Medial rotation	<i>r<sup>s</sup></i> = -0.36
AROM Lateral rotation	<i>r<sup>s</sup></i> = -0.26
Perceived recovery (anchor)	<i>r<sup>s</sup></i> = -0.61

DASH disabilities of the arm, shoulder and hand questionnaire, NPRS numeric pain rating scale, SPADI shoulder pain and disability index, SF-36 The Short Form 36 Health Survey (SF-36), AROM active range of motion

**Smallest detectable change**

The mean time period between T1 and T2 was 7.4 days (SD, 2.1). The ICC of the DASH was found to be 0.91. The SEM was 4.3, and the SDC was found to be 11.8.

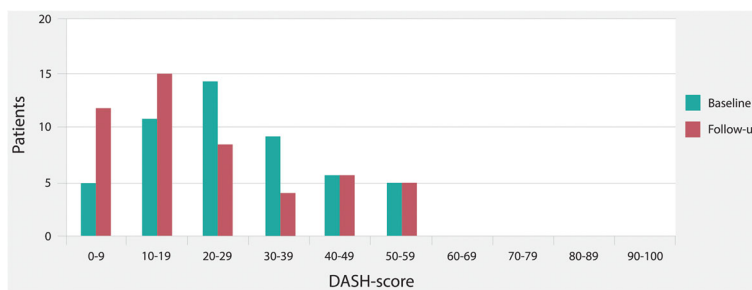
**Minimal important change**

Patients who reported deterioration on the perceived recovery (anchor) were excluded from the calculation of MIC. The anchor was considered appropriate since the correlation of both instruments with the anchor was higher than 0.5 (Table 4).

The MIC for the total change scores of DASH was 4.4, with a sensitivity of 0.77 and specificity of 0.69 (Table 5). Figure 3 illustrates the anchor-based distribution of the change scores for the improved and unchanged group, and the MIC value. Using the MIC as a cut-off point, 23.4% of the patients who were “improved” according to the anchor, had a lower change score and were considered false negatives. 30.8% of the patients, who were “unchanged” according to the anchor, had a higher change score and were considered false positives.

MIC values for low and high baseline scores was respectively 4.4 and 7.7. The median DASH score was 26.4. Twenty-two patients (15 improved, 7 unchanged) had low DASH baseline scores (i.e. DASH ≤26.4). Twenty-one patients (15 improved, 6 unchanged) had





**Fig. 2** Score distribution of the DASH before and after the physical therapy. Higher values indicate greater disability

high DASH baseline scores (DASH > 26.4). The MIC data are presented in Table 5.

**Discussion**

The DASH is often used in research and clinical practice to evaluate disability and the effect of interventions for shoulder disorders. Therefore, it is important to evaluate its ability to capture change, and especially the smallest change score that patients perceive as important.

The study of responsiveness was conducted following all requirements of the COSMIN-checklist [41]: (1) Reporting the percentage of missing items, (2) describing how missing items were handled, (3) appropriate sample size. The total sample size included in the analysis, was within the range described as good ( $n = 50-99$ ), except for the subgroup analysis and the calculation of the MIC which was poor ( $n < 30$ ). (4) A longitudinal design with three measurements was used, (5) the time interval was stated, (6) the intervention was described, and (7) the proportion of patients that changed (i.e. improved, unchanged and deteriorated) was described. Regarding the hypotheses testing, (8) the hypotheses were formulated a priori, (9, 10) the expected direction and magnitude of correlations of the change scores were defined, (11, 12) the comparator instruments and its measurement properties were described. (13, 14, 16) The design, method and statistics are considered appropriate, (15) the criterion for change (anchor) can be considered to be a reasonable gold standard according to its correlation with the DASH (>0.50), (17) both correlation between the change scores and ROC curve were calculated, and (18) sensitivity and

specificity for the patients who were ‘improved’ and ‘unchanged’ was determined.

For evaluating the responsiveness, both a criterion and a construct approach was conducted according to the COSMIN recommendations [9]. Using the criterion approach, this study found evidence for good responsiveness of DASH (AUC = 0.77, CI: 0.63, 0.92). This indicates that the questionnaire was able to distinguish patients who reported to have improved, from those who remained unchanged. The AUC-value found in the present study is similar to the findings reported in recent studies. In patients with a variety of shoulder disorders, Lundquist et al. [42] found an AUC of 0.76 (0.62–0.90). Michener et al. [43] found the AUC to be 0.79 (0.69–0.89) in patients with shoulder impingement. Similar to our findings, no floor or ceiling effect was reported in these studies, or in studies that included patients with upper-limb disorders [44, 45].

According to the construct approach, a total of 6 hypotheses were a priori formulated regarding the relationships between changes on the DASH and changes on related measures. None of the hypotheses were refuted. DASH showed little to moderate correlation with impairment measures, as is consistent with findings in other studies [46, 47].

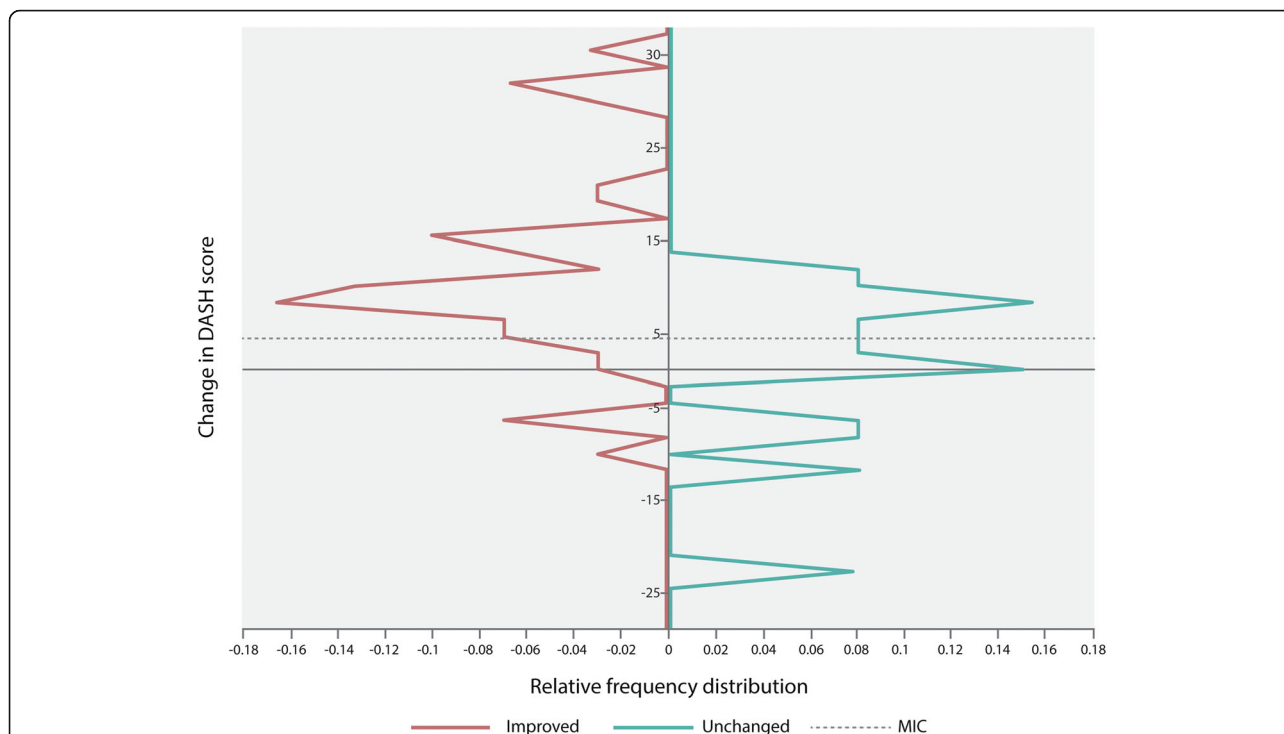
We found a MIC value of 4.4 for the total DASH data. This implies that a change of five on the scale is likely to be considered important by individual patients. However, measurement error by SDC was found to be 11.8 which is a larger value than all calculated MIC values. This means that individual patients may be considered a MIC value to be an important change. However, as it cannot be distinguished from measurement error it is not statistically significant [48]. Measurement error has to be taken into consideration when interpreting change scores. We should both be confident that the change in DASH is not simply due to measurement error and that it is sufficiently high to be important to individual patients. Hence, a change value must be over 11.8 for the DASH to be considered as an important change. In the present study 11 patients had a change value above

**Table 5** Minimal important changes of the DASH ( $n = 50$ )

	Total sample			Low baseline scores			High baseline scores		
	MIC	SN	SP	MIC	SN	SP	MIC	SN	SP
DASH	4.4	0.77	0.69	4.4	0.73	0.71	7.7	0.73	0.83

Median DASH score = 26.4

DASH disabilities of the arm, shoulder and hand questionnaire, MIC minimal important change, SN sensitivity, SP specificity



**Fig. 3** Anchor-based MIC distribution of the DASH with indication of the ROC cut-off point, after dichotomizing the patients in groups of improved and unchanged. MIC 4.4. Sensitivity at this point was 0.77 and specificity was 0.69

11.8, although 30 patients were considered improved according to the anchor.

Additionally, we found a higher MIC (7.7) for the patients with high baseline scores (Table 5). This supports previous findings that patients with higher baseline scores needs greater changes to be considered as important [37, 40]. MIC can also be expressed as percentage of baseline values [40]. When calculating the MIC values in our study as percentage, the MIC for the total DASH is 15.8% (4.4 of 27.9), and for the subgroups with low and high baseline scores, 25.9% (4.4 of 17) and 19.6% (7.7 of 39.2), respectively.

The SDC value found in our study is comparable with SDC values reported by others [36, 49, 50]. The MIC value (4.4) in the present study is, however, much lower as compared to results found in the literature. Gummesson et al. [51] found a MIC of 10 and van Kampen et al. [36] a MIC of 12.4. Beaton et al. [49] studied heterogeneous shoulder patients and compared different approaches for determine the MIC. Their results ranged from 3.9 to 15. These differing findings indicate how dependent the MIC is on the anchor. The anchor used in this study was a perceived recovery recorded on a 3-level ordinal scale, similar to what was used in a recent study [47] on responsiveness of the DASH. Using a 7- or 15-point Likert scale could have been better suited for this study, resulting in a

more refined dichotomising between the subgroups. Nevertheless, such anchors have been criticized for not being capable of correlating functional change across varying lengths of time [52]. The anchor used in this present study can be questioned due to the large overlap between the distributions of the improved and unchanged group of patients illustrated in Fig. 3 according to the MIC value. However, in this study the anchor correlated above 0.5 with the change score of the DASH, and is then proposed to be a valid anchor according to a recent study [40]. Since there is no clear agreement of which method is the best to examine MIC, other approaches could have resulted in different MIC values. In addition, both the anchor-based approach and the COSMIN-group’s definitions of MIC and SDC used in this study, leads to a clear distinction between these two concepts [9, 53].

The present study is, to our knowledge, the first study to explicitly follow the COSMIN panel recommendation regarding examination of responsiveness of the DASH. However, there are some limitations as to how the findings should be interpreted. First, the COSMIN-checklist was developed to improve the selection of measurement instruments and how to evaluate these measurement properties. However, the checklist has not yet been tested for its reliability, and according to the authors it needs further refinement [41]. Second, the total sample

size of this study is moderate ( $n = 50$ ). Although it is within the COSMIN recommendation for examination of responsiveness, the sample size of the unchanged subgroup ( $n = 13$ ) used for the analyses was small. Third, an ordinal scale with more scoring options might have been preferred.

## Conclusions

In conclusion, the Norwegian version of the DASH was found to be satisfactory responsive when applied on patients with subacromial pain. Based on the SDC and MIC, a change value on the DASH must be above 11.8 to be considered as an important change that is not due to measurement error.

## Abbreviations

AROM: Active range of motion; AUC: Area under the ROC curve; COSMIN: COnsensus-based Standards for the selection of health Measurement Instruments; DASH: The disabilities of the arm, shoulder and hand questionnaire; MIC: Minimal important change; NPRS: Numeric pain rating scale; PROM: Patient reported outcome measure; ROC: Receiver operating characteristic; SD: Standard deviation; SDC: Smallest detectable change; SEM: Standard error of measurement; SF-36: Short Form 36 Health Survey; SPADI: Shoulder pain and disability index; SPS: Subacromial pain syndrome

## Acknowledgements

We would like to thank all the patients who participated in this study.

## Funding

This work was supported by the Norwegian Fund for Post-Graduate Training in Physiotherapy (Grant numbers 76,326, 2015).

## Availability of data and materials

The data will not be share due to personal data protection.

## Authors' contributions

TR carried out the design of the study, conducted data analysis, and prepared the manuscript. BH and IS gathered data. YR, BH, IS and LIS revised the manuscript critically. LIS supervised the project. All authors read and approved the final manuscript.

## Competing interests

The authors declare that they have no competing interests.

## Consent for publication

Not applicable.

## Ethics approval and consent to participate

Informed consent was obtained from all participants. The study protocol was approved by the Regional committees for medical and health research ethics (REC South East), and was carried out in accordance with the Helsinki Declaration. Date: 5th June 2008 ID: 163-08086d 1.2008.131.

## Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

## Author details

<sup>1</sup>Faculty of Health Sciences, Oslo and Akershus University College of Applied Sciences, Postboks 4, St Olavs Plass, Oslo, Norway. <sup>2</sup>Department of Physiotherapy, Martina Hansens Hospital, Bærum, Norway. <sup>3</sup>Norwegian Research Center for Active Rehabilitation, Department of Orthopaedics, Oslo University Hospital, Oslo, Norway. <sup>4</sup>Department of Global Public Health and Primary Care, University of Bergen, Bergen, Norway.

Received: 28 June 2016 Accepted: 2 June 2017

Published online: 08 June 2017

## References

- Luime JJ, Koes BW, Hendriksen IJM, Burdorf A, Verhagen AP, Miedema HS, et al. Prevalence and incidence of shoulder pain in the general population; a systematic review. *Scand J Rheumatol*. 2004;33:73–81.
- Michener LA, Walsworth MK, Burnet EN. Effectiveness of rehabilitation for patients with subacromial impingement syndrome: a systematic review. *J Hand Ther*. 2004;17:152–64.
- Fayad F, Mace Y, Lefevre-Colau MM. Shoulder disability questionnaires: a systematic review. *Ann Readapt Med Phys*. 2005;48:298–306.
- Snyder CF, Aaronson NK, Choucair AK, Elliott TE, Greenhalgh J, Halyard MY, et al. Implementing patient-reported outcomes assessment in clinical practice: a review of the options and considerations. *Qual Life Res*. 2012;21:1305–14.
- Valderas JM, Kotzeva A, Espallargues M, Guyatt G, Ferrans CE, Halyard MY, et al. The impact of measuring patient-reported outcomes in clinical practice: a systematic review of the literature. *Qual Life Res*. 2008;17(2):179–93.
- Terwee CB, Bot SD, de Boer MR, van der Windt DA, Knol DL, Dekker J, et al. Quality criteria were proposed for measurement properties of health status questionnaires. *J Clin Epidemiol*. 2007;60:34–42.
- Mokkink LB, Terwee CB, Patrick DL, Alonso J, Stratford PW, Knol DL, et al. The COSMIN study reached international consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes. *J Clin Epidemiol*. 2010;63:737–45.
- Mokkink LB, Prinsen CAC, Bouter LM, de Vet HCW, Terwee CB. The COnsensus-based standards for the selection of health measurement INstruments (COSMIN) and how to select an outcome measurement instrument. *Braz J Phys Ther*. 2016;20(2):105–13. doi:10.1590/bjpt-rbf.2014.0143.
- de Vet HCW, Knol DL, Terwee CB, Mokkink LB. *Measurement in medicine*. Cambridge: Cambridge University Press; 2011.
- Roy JS, Macdermid JC, Woodhouse LJ. Measuring shoulder function: a systematic review of four questionnaires. *Arthritis Rheum*. 2009;61:623–32.
- Hudak PL, Amadio PC, Bombardier C. Development of an upper extremity outcome measure: the DASH (disabilities of the arm, shoulder, and hand). *Am J Ind Med*. 1996;29:602–8.
- Finsen V. Norwegian version of the DASH questionnaire for examination of the arm shoulders and hand. *Tidsskr Nor Laegeforen*. 2008;128:1070.
- Beaton DE, Katz JN, Fossel AH, Wright JG, Tarasuk V, Bombardier C. Measuring the whole or the parts? Validity, reliability, and responsiveness of the disabilities of the arm, shoulder and hand outcome measure in different regions of the upper extremity. *J Hand Ther*. 2001;14:128–46.
- Angst F. The new COSMIN guidelines confront traditional concepts of responsiveness. *BMC Med Res Methodol*. 2011;14:152.
- Revicki D, Hays RD, Cella D, Sloan J. Recommended methods for determining responsiveness and minimally important differences for patient-reported outcomes. *J Clin Epidemiol*. 2008;61:102–9.
- 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. *Semin Arthritis Rheum*. 2003;33:168–84.
- Solway S, Beaton DE, McConnell S, Bombardier C. The DASH outcome measure User's manual: disabilities of the arm, shoulder and hand. *Inst. For work & health*; 2002.
- Haldorsen B, Svege I, Roe Y, Bergland A. Reliability and validity of the Norwegian version of the disabilities of the arm, shoulder and hand questionnaire in patients with shoulder impingement syndrome. *BMC Musculoskelet Disord*. 2014;15:78.
- Roach KE, Budiman-Mak E, Songsiridej N, Lertratanakul Y. Development of a shoulder pain and disability index. Volume 4; 1991.
- Bot SDM, 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;63:335–41.
- Hill CL, Lester S, Taylor AW, Shanahan ME, Gill TK. Factor structure and validity of the shoulder pain and disability index in a population-based study of people with shoulder symptoms. *BMC Musculoskelet Disord*. 2011;12:8.



22. Desai AS, Dramis A, Hearnden AJ. Critical appraisal of subjective outcome measures used in assessment of shoulder disability. *Ann R Coll Surg Engl*. 2010;92:9–13.
23. Ekeberg OM, Bautz-Holter E, Tveitå EK, Keller A, Juel NG, Brox JI. Agreement, reliability and validity in 3 shoulder questionnaires in patients with rotator cuff disease. *BMC Musculoskelet Disord*. 2008;9:68.
24. Largacha M, Parsons IM, Campbell B, Titelman RM, Smith KL, Matsen F. Deficits in shoulder function and general health associated with sixteen common shoulder diagnoses: a study of 2674 patients. *J Shoulder Elb Surg*. 2006;15:30–9.
25. Ware JE, Snow KK, Kosinski M, Gandek B. SF-36 health Survey manual and interpretation guide. 1993.
26. Loge JH, Kaasa S, Hjermstad MJ, Kvien TK. Translation and performance of the Norwegian SF-36 health Survey in patients with rheumatoid arthritis. I. Data quality, scaling assumptions, reliability, and construct validity. *J Clin Epidemiol*. 1998;51:1069–76.
27. Mintken PE, Glynn P, Cleland JA. Psychometric properties of the shortened disabilities of the arm, shoulder, and hand questionnaire (QuickDASH) and numeric pain rating scale in patients with shoulder pain. *J Shoulder Elb Surg*. 2009;18:920–6.
28. Michener LA, Snyder AR, Leggin BG. Responsiveness of the numeric pain rating scale in patients with shoulder pain and the effect of surgical status. *J Sport Rehabil*. 2011;20:115–28.
29. Kamper SJ, Maher CG, Mackay G. Global rating of change scales: a review of strengths and weaknesses and considerations for design. *J Man Manip Ther*. 2009;17:163–70.
30. Kamper SJ, Ostelo RWJG, Knol DL, Maher CG, de Vet HCW, Hancock MJ. Global perceived effect scales provided reliable assessments of health transition in people with musculoskeletal disorders, but ratings are strongly influenced by current status. *J Clin Epidemiol*. 2010;63:760–6.
31. Guyatt GH, Norman GR, Juniper EF, Griffith LE. A critical look at transition ratings. *J Clin Epidemiol*. 2002;55:900–8.
32. Schmitt J, Di Fabio RP. The validity of prospective and retrospective global change criterion measures. *Arch Phys Med Rehabil*. 2005;86:2270–6.
33. Munro BH. *Statistical methods for health care research*. 5th ed. Philadelphia: Lippincott Williams & Wilkins; 2005.
34. de Boer MR, Moll AC, de Vet HCW, Terwee CB, Völker-Dieben HJM, van Rens GHMB. Psychometric properties of vision-related quality of life questionnaires: a systematic review. *Ophthalmic Physiol Opt*. 2004;24:257–73.
35. Weir JP. Quantifying test-retest reliability using the intraclass correlation coefficient and the SEM. *J Strength Cond Res*. 2005;19:231–40.
36. van Kampen DA, Willems WJJ, van Beers LWAH, Castelein RM, Scholtes VAB, Terwee CB. Determination and comparison of the smallest detectable change (SDC) and the minimal important change (MIC) of four-shoulder patient-reported outcome measures (PROMs). *J Orthop Surg Res*. 2013;8:40.
37. de Vet HCW, Ostelo RWJG, Terwee CB, van der Roer N, Knol DL, Beckerman H, et al. Minimally important change determined by a visual method integrating an anchor-based and a distribution-based approach. *Qual Life Res*. 2006;16:131–42.
38. Cella D, Hahn EA, Dineen K. Meaningful change in cancer-specific quality of life scores: differences between improvement and worsening. *Qual Life Res*. 2002;11:207–21.
39. Schuller W, Ostelo RWJG, Janssen R, de Vet HCW. The influence of study population and definition of improvement on the smallest detectable change and the minimal important change of the neck disability index. *Health Qual Life Outcomes*. 2014;12:53.
40. De Vet HCW, Foumani M, Scholten MA, Jacobs WCH, Stiggelbout AM, Knol DL, et al. Minimally important change values of a measurement instrument depend more on baseline values than on the type of intervention. *J Clin Epidemiol*. 2015;68:518–24.
41. 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.
42. Lundquist CB, Døssing K, Christiansen DH. Responsiveness of a Danish version of the disabilities of the arm, shoulder and hand (DASH) questionnaire. *Dan Med J*. 2014;61:A4813.
43. Michener LA, Snyder Valier AR, McClure PW. Defining substantial clinical benefit for patient-rated outcome tools for shoulder impingement syndrome. *Arch Phys Med Rehabil*. 2013;94:725–30.
44. Franchignoni F, Vercelli S, Giordano A, Sartorio F, Bravini E, Ferriero G. Minimal clinically important difference of the disabilities of the arm, shoulder and hand outcome measure (DASH) and its shortened version (QuickDASH). *J Orthop Sports Phys Ther*. 2014;44:30–9.
45. Lehman LA, Sindhu BS, Shechtman O, Romero S, Velozo CA. A comparison of the ability of two upper extremity assessments to measure change in function. *J Hand Ther*. 2010;23:31–40.
46. Offenbächer M, Ewert T, Sangha O, Stucki G. Validation of a German version of the "disabilities of arm, shoulder and hand" questionnaire (DASH-G). *Z Rheumatol*. 2003;62:168–77.
47. Fayad F, Lefevre-Colau MM, Mace Y, Gautheron V, Fermanian J, Roren A, et al. Responsiveness of the French version of the disability of the arm, shoulder and hand questionnaire (F-DASH) in patients with orthopaedic and medical shoulder disorders. *Jt Bone Spine*. 2008;75:579–84.
48. Terwee CB, Roorda LD, Dekker J, Bierma-Zeinstra SM, Peat G, Jordan KP, et al. Mind the MIC: large variation among populations and methods. *J Clin Epidemiol*. 2010;63:524–34.
49. Beaton DE, Van Eerd D, Smith P, Van Der Velde G, Cullen K, Kennedy CA, et al. Minimal change is sensitive, less specific to recovery: a diagnostic testing approach to interpretability. *J Clin Epidemiol*. 2011;64:487–96.
50. Schmitt JS, Di Fabio RP. Reliable change and minimum important difference (MID) proportions facilitated group responsiveness comparisons using individual threshold criteria. *J Clin Epidemiol*. 2004;57:1008–18.
51. Gummesson C, Atroshi I, Ekdhah C. The disabilities of the arm, shoulder and hand (DASH) outcome questionnaire: longitudinal construct validity and measuring self-rated health change after surgery. *BMC Musculoskelet Disord*. 2003;4:11.
52. Schmitt J, Abbott JH. Global ratings of change do not accurately reflect functional change over time in clinical practice. *J Orthop Sports Phys Ther*. 2015;45:106–11.
53. de Vet HCW, Terluin B, Knol DL, Roorda LD, Mokkink LB, Ostelo RWJG, et al. Three ways to quantify uncertainty in individually applied "minimally important change" values. *J Clin Epidemiol*. 2010;63:37–45.
54. Raven EEJ, Haverkamp D, Siersevelt IN, Van Montfoort DO, Pöhl RG, Blankevoort L, et al. Construct validity and reliability of the disability of arm, shoulder and hand questionnaire for upper extremity complaints in rheumatoid arthritis. *J Rheumatol*. 2008;35:2334–8.
55. Ebersole GC, Davidge K, Damiano M, Mackinnon SE. Validity and responsiveness of the DASH questionnaire as an outcome measure following ulnar nerve transposition for cubital tunnel syndrome. *Plast Reconstr Surg*. 2013;132:81e–90e.
56. Angst F, Pap G, Mannion AF, Herren DB, Aeschlimann A, Schwyzer HK, et al. Comprehensive assessment of clinical outcome and quality of life after total shoulder arthroplasty: usefulness and validity of subjective outcome measures. *Arthritis Care Res*. 2004;51:819–28.
57. Angst F, Schwyzer H-K, Aeschlimann A, Simmen BR, Goldhahn J. Measures of adult shoulder function: disabilities of the arm, shoulder, and hand questionnaire (DASH) and its short version (QuickDASH), shoulder pain and disability index (SPADI), American shoulder and elbow surgeons (ASES) society standardized shoulder. *Arthritis Care Res (Hoboken)*. 2011;63(Suppl 1):S174–88.
58. Macdermid JC, Khadilkar L, Birmingham TB, Athwal GS. Validity of the QuickDASH in patients with shoulder-related disorders undergoing surgery. *J Orthop Sports Phys Ther*. 2015;45:25–36.
59. Dierckx R, Bron C, Dorrestijn O, Meskers C, Naber R, de Ruiter T, et al. Guideline for diagnosis and treatment of subacromial pain syndrome. *Acta Orthop*. 2014;85:314–22.