# Psychometric properties of the shortened disabilities of the arm, shoulder, and hand questionnaire (QuickDash) and numeric pain rating scale in patients with shoulder pain

## Introduction

Function of the shoulder has been conventionally assessed by objective measures such as ROM and strength. However, although these are useful they have no direct clinical meaning to patients who just want to be free of pain and able to perform their daily activities. Nowadays, the efficacy of treatment is more often evaluated with subjective measures that assess the ability to function in daily life.

It is important to investigate the clinimetric quality of an outcome measure to determine whether or not it is useful in the clinical setting. The quality of an O.M. is shown by its reliablilty, validity and responsiveness (i.e. MCID).

While the QuickDash has demonstrated both reliability and validity there are no studies to date investigating the MCID. Also while the QuickDash has been used in broad populations of patients with upper extremity pain, its performance in patients with shoulder pain has not been investigated.

In addition the MCID of the Numeric Pain Rating Scale (NPRS) has been reported in broad populations as 2 points however this has not been investigated in patients whose primary complaint is shoulder pain.

## Aims & Objectives

The aim of this study was to investigate the psychometric properties of the QuickDash and the NPRS in patients whose primary complaint was shoulder pain.

## Methodology

The design of this study was a single-group repeated measures design thus there was no blinding, no random allocation and a relatively small number of subjects was needed.

Data was collected from 101 consecutive patients with a primary complaint of shoulder pain who presented to one of two outpatient orthopedic physical therapy clinics in Colorado and Newton, MA.

Inclusion criteria were aged over 18 and able to read and understand English. Demographic information was collected from each subject. They also completed a QuickDash, NPRS and a body diagram at baseline.

All subjects then received physical therapy as prescribed by the treating therapist. The QuickDash, NPRS, body diagram were repeated along with the Global Rating of Change (GRC). This re-evaluation was usually 2-4 weeks after the baseline assessment.

The GRC is a 15-point scale completed by the patient, which ranges from -7 (a very great deal worse) to 0 (about the same) to +7 (a very great deal better). It has been used to identify responsiveness and the MCID for other O.M. in cases where there is no gold standard available.

For the purpose of data analysis subjects were divided into two groups; stable (GRC between -3 to +3, i.e. minimal to no change) and those with a clinically important improvement (GRC >+3).

Baseline variables were analysed using chi-squared tests and t-tests.

Test retest reliability of the Q.D. and the NPRS was measured using ICC with 95% CI.

Construct validity of the Q.D. and NPRS were examined by comparing the change in scores for the stable and improved groups using ANOVA analysis for the repeated measures at baseline and re-evaluation. The authors hypothesized that if a patient was 'stable' their Q.D. and NPRS score would not change whereas 'improved' patients would demonstrate significant changes in these scores.

Responsiveness of the Q.D. was measured by two methods. First, by constructing ROC curves to calculate the AUC for both the QD and NPRS, .82 and .67 respectively. And second, by correlating the change scores of the QD to the GRC and change scores for the NPRS.

Minimum detectable change (MDC) was calculated by determining the standard error of measure (SEM) for the QD for the stable group. The 90% CI of this was calculated.

MCID was calculated by identifying the point on the ROC curve nearest the upper left-hand corner, this is considered to be the best cut-off for distinguishing improved and stable patients (highest proportion of true positives to false positives).

**Results**

Out of the 101 subjects studied, 22 remained stable and 79 improved. A significant number of the group that improved had sustained a traumatic injury rather than a non-traumatic injury.

The authors showed that the QD is a reliable OM by indicating test-retest reliability with ICC values of 0.90 (95% CI). While the NPRS showed only moderate agreement when test was reproduced.

The pre-test and post-test scores of the QD differed for the stable group versus the improved group, thus confirming the construct validity of the QD.

The pre and post-test scores of the NPRS also differed for the stable and improved groups based on the GRC.

The AUC for the QD was 0.82 which indicates high diagnostic accuracy whereas the AUC for the NPRS was 0.67 which indicates poor/ fair accuracy.

The MCID of the QD was calculated as 8% points with a high sensitivity and specificity. The MCID of the NPRS was 1.1 with only fair sensitivity and specificity.

**Discussion**

The study shows that the QD exhibits excellent test retest reliability. This is substantiated by previous research. The NPRS exhibits moderate test retest reliability when used on a group of patients with shoulder pain.

It also demonstrates that both the QD and the NPRS have significant construct validity, i.e. the QD showed a reduction in disability among patients rating themselves as improved versus stable.  This also substantiates the findings of previous studies.

The two methods used by the authors to investigate the responsiveness of the QD and the NPRS showed a moderate but significant correlation for the QD and a small but significant correlation for the NPRS.

The study determined an MCID of 1.1 points for the NPRS which is consistent with findings of previous work. It determined the MCID of 8% points for the QD which is equivalent to less than a 4 point change out of 55 in the summation of the questionnaire. This indicates that the QD is indeed responsive to change in patients with shoulder pain.

The use of the GRC as opposed to a gold standard could be considered a limitation to this study. The criterion validity could be compromised due to lack of comparison with a gold standard. However the authors allude to this limitation and it appears it is unavoidable until further research is conducted to produce a gold standard measure or a standardized approach to examining psychometric properties of OM.

Also although the test retest reliability was assessed correctly, the internal reliability (i.e. correlation of subscales) was not mentioned.

Floor and ceiling effects of the measures were not tested thus these OM may fail to detect a worse score in patients clinically deteriorated. In fact patients that disimproved were excluded from the results so we have no information as to how the QD can detect a clinical deterioration.

**Conclusion**
This study investigated patients that presented to an out-patient setting therefore the reliability, validity, responsiveness and MCID calculated can only be attributed to this patients in this setting. It is not known whether these results can be held up for patients that present for treatment in for example a hospital setting.