Reliability, validity, responsiveness, and minimal important change of the Disabilities of the Arm, Shoulder and Hand and Constant-Murley scores in patients with a humeral shaft fracture

Kiran C. Mahabier, MD, Dennis Den Hartog, MD PhD, Nina Theyskens, Michael H.J. Verhofstad, MD PhD, Esther M.M. Van Lieshout, MSc PhD*, on behalf of the HUMMER Trial Investigators Collaboration

Trauma Research Unit Department of Surgery, Erasmus MC, University Medical Center Rotterdam, Rotterdam, The Netherlands

*** Corresponding author:**

Esther M.M. Van Lieshout, MSc PhD

Erasmus MC, University Medical Center Rotterdam

Trauma Research Unit Department of Surgery

P.O. Box 2040, 3000 CA Rotterdam, The Netherlands

Mail: e.vanlieshout@erasmusmc.nl

†A list of the Hummer Investigators Collaboration is available in the Appendix.

P. Koen Bos, MD, PhD1, Maarten W.G.A. Bronkhorst, MD2,

Milko M.M. Bruijninckx, MD3, Jeroen De Haan, MD, PhD4, P. Ted Den Hoed, MD, PhD5,

Martin G. Eversdijk, MD6, J. Carel Goslings, MD, PhD7, Robert Haverlag, MD8,

Martin J. Heetveld, MD, PhD9, Albert J.H. Kerver, MD, PhD10, Karel A. Kolkman, MD11,

Peter A. Leenhouts, MD12, Sven A.G. Meylaerts, MD, PhD2, Ron Onstenk, MD13,

Martijn Poeze, MD, PhD14, Rudolf W. Poolman, MD, PhD15, Bas J. Punt, MD16,

Ewan D. Ritchie, MD17, W. Herbert Roerdink, MD, PhD18, Gert R. Roukema, MD19,

Jan Bernard Sintenie, MD20, Nicolaj M.R. Soesman, MD21,

Maarten Van der Elst, MD, PhD22, Frank H.W.M. Van der Heijden, MD, PhD23,

Frits M. Van der Linden, MD24, Peer Van der Zwaal, MD, PhD25, Jan P. Van Dijk, MD26,

Hans-Peter W. Van Jonbergen, MD, PhD27, Egbert J.M.M. Verleisdonk, MD, PhD28,

Jos P.A.M. Vroemen, MD, PhD29, Marco Waleboer, MD30, Philippe Wittich, MD, PhD31,

Wietse P. Zuidema, MD32

1 Department of Orthopaedic Surgery, Erasmus MC, University Medical Center Rotterdam, Rotterdam, The Netherlands

2 Department of Surgery, MC Haaglanden Bronovo, The Hague, The Netherlands

3 Department of Surgery, IJsselland Hospital, Capelle a/d Ijssel, The Netherlands

4 Department of Surgery, Westfriesgasthuis, Hoorn, The Netherlands

5 Department of Surgery, Ikazia Hospital, Rotterdam, The Netherlands

6 Department of Surgery, St. Jansdal Hospital, Harderwijk, The Netherlands

7 Trauma Unit Department of Surgery, Academic Medical Center, Amsterdam, The Netherlands

8 Department of Surgery, Onze Lieve Vrouwe Gasthuis, Amsterdam, The Netherlands

9 Department of Surgery, Spaarne Gasthuis, Haarlem, The Netherlands

10 Department of Surgery, Franciscus Gasthuis & Vlietland, Rotterdam, The Netherlands

11 Department of Surgery, Rijnstate, Arnhem, The Netherlands

12 Zaans Medical Center, Zaandam, The Netherlands

13 Department of Orthopaedic Surgery, Groene Hart Hospital, Gouda, The Netherlands

14 Department of Trauma Surgery, Maastricht University Medical Center, Maastricht, The Netherlands

15 Department of Orthopaedic Surgery, Onze Lieve Vrouwe Gasthuis, Amsterdam, The Netherlands

16 Department of Surgery, Albert Schweitzer Hospital, Dordrecht, The Netherlands

17 Department of Surgery, Alrijne Hospital, Leiderdorp, The Netherlands

18 Department of Surgery, Deventer Hospital, Deventer, The Netherlands

19 Department of Surgery, Maasstad Hospital, Rotterdam, The Netherlands

20 Department of Surgery, Elkerliek Hospital, Helmond, The Netherlands

21 Department of Surgery, Franciscus Gasthuis & Vlietland, Schiedam, The Netherlands

22 Department of Surgery, Reinier de Graaf Gasthuis, Delft, The Netherlands

23 Department of Surgery, St Elisabeth Hospital, Tilburg, The Netherlands

24 Department of Surgery, Groene Hart Hospital, Gouda, The Netherlands

25 Department of Orthopaedic Surgery, MC Haaglanden Bronovo, The Hague, The Netherlands

26 Department of Surgery, Hospital Gelderse Vallei, Ede, The Netherlands

27 Department of Orthopaedic Surgery, Deventer Hospital, Deventer, The Netherlands

28 Department of Surgery, Diakonessenhuis, Utrecht, The Netherlands

29 Department of Surgery, Amphia Hospital, Breda, The Netherlands

30 Department of Surgery, Admiraal De Ruyter Hospital, Vlissingen, The Netherlands

31 Department of Surgery, St. Antonius Hospital, Nieuwegein, The Netherlands

32 Department of Trauma Surgery, VU University Medical Center, Amsterdam, The Netherlands

Appendix

HUMMER Trial investigators (all from The Netherlands)

P. Koen Bos, MD PhD, Department of Orthopaedic Surgery, Erasmus MC, University Medical Center Rotterdam

Maarten W.G.A. Bronkhorst, MD, Department of Surgery, MC Haaglanden Bronovo, The Hague

Milko M.M. Bruijninckx, MD, Department of Surgery, IJsselland Hospital, Capelle a/d Ijssel

Jeroen De Haan, MD PhD, Department of Surgery, Westfriesgasthuis, Hoorn

P. Ted Den Hoed, MD PhD, Department of Surgery, Ikazia Hospital, Rotterdam

Martin G. Eversdijk, MD, Department of Surgery, St. Jansdal Hospital, Harderwijk

J. Carel Goslings, MD PhD, Trauma Unit Department of Surgery, Academic Medical Center, Amsterdam

Robert Haverlag, MD, Department of Surgery, Onze Lieve Vrouwe Gasthuis, Amsterdam

Martin J. Heetveld, MD PhD, Department of Surgery, Spaarne Gasthuis, Haarlem

Albert J.H. Kerver, MD PhD, Department of Surgery, Franciscus Gasthuis & Vlietland, Rotterdam

Karel A. Kolkman, MD, Department of Surgery, Rijnstate, Arnhem

Peter A. Leenhouts, MD, Zaans Medical Center, Zaandam

Sven A.G. Meylaerts, MD PhD, Department of Surgery, MC Haaglanden Bronovo, The Hague

Ron Onstenk, MD, Department of Orthopaedic Surgery, Groene Hart Hospital, Gouda

Martijn Poeze, MD PhD, Department of Trauma Surgery, Maastricht University Medical Center

Rudolf W. Poolman, MD PhD, Department of Orthopaedic Surgery, Onze Lieve Vrouwe Gasthuis, Amsterdam

Bas J. Punt, MD, Department of Surgery, Albert Schweitzer Hospital, Dordrecht

Ewan D. Ritchie, MD, Department of Surgery, Alrijne Hospital, Leiderdorp

W. Herbert Roerdink, MD PhD, Department of Surgery, Deventer Hospital

Gert R. Roukema, MD, Department of Surgery, Maasstad Hospital, Rotterdam

Jan Bernard Sintenie, MD, Department of Surgery, Elkerliek Hospital, Helmond

Nicolaj M.R. Soesman, MD, Department of Surgery, Franciscus Gasthuis & Vlietland, Schiedam

Maarten Van der Elst, MD PhD, Department of Surgery, Reinier de Graaf Gasthuis, Delft

Frank H.W.M. Van der Heijden, MD PhD, Department of Surgery, St Elisabeth Hospital, Tilburg

Frits M. Van der Linden, MD, Department of Surgery, Groene Hart Hospital, Gouda

Peer Van der Zwaal, MD PhD, Department of Orthopaedic Surgery, MC Haaglanden Bronovo, The Hague

Jan P. Van Dijk, MD, Department of Surgery, Hospital Gelderse Vallei, Ede

Hans-Peter W. Van Jonbergen, MD PhD, Department of Orthopaedic Surgery, Deventer Hospital

Egbert J.M.M. Verleisdonk, MD PhD, Department of Surgery, Diakonessenhuis, Utrecht

Jos P.A.M. Vroemen, MD PhD, Department of Surgery, Amphia Hospital, Breda

Marco Waleboer, MD, Department of Surgery, Admiraal De Ruyter Hospital, Vlissingen

Philippe Wittich, MD PhD, Department of Surgery, St. Antonius Hospital, Nieuwegein

Wietse P. Zuidema, MD, Department of Trauma Surgery, VU University Medical Center, Amsterdam

Disclaimer

This work was funded by a grant from the Osteosynthesis and Trauma Care (OTC) Foundation (reference number 2013-DHEL).

The authors, their immediate families, and any research foundations with which they are affiliated have not received any financial payments or other benefits from any commercial entity related to the subject of this article.

Ethical approval

This study was exempted by the Erasmus MC Medical Research Ethics Committee (No. MEC-2015-217).

ACKNOWLEDGMENTS

Dr. Wim E. Tuinebreijer (clinical epidemiologist, Trauma Research Unit, Department of Surgery, Erasmus MC, University Medical Center Rotterdam, Rotterdam, The Netherlands) is acknowledged for statistical advice on the MIC analysis. Tim Van der Torre, Jelle E. Bousema, Boyd C. Van der Schaaf, Joyce Van Veldhuizen, Marije C.A.W. Notenboom, Yordi Claes, and Boudijn S.H. Joling (medical students, Trauma Research Unit, Department of Surgery, Erasmus MC, University Medical Center Rotterdam, Rotterdam, The Netherlands) are acknowledged for their assistance in data collection.

ABSTRACT

Background: The Disabilities of the Arm, Shoulder and Hand (DASH) and Constant-Murley scores are commonly used instruments. The DASH is patient reported, and the Constant-Murley combines a clinician reported and a patient-reported part. For patients with a humeral shaft fracture, their validity, reliability, responsiveness, and Minimal Important Change (MIC) have not been published. This study evaluated the measurement properties of these instruments in patients who sustained a humeral shaft fracture.

Methods: The DASH and Constant-Murley instruments were completed five times until one year after trauma. Pain score, Short Form 36, and EuroQol-5D were completed for comparison. Internal consistency was determined by the Cronbach α. Construct and longitudinal validity were evaluated by assessing hypotheses about expected Spearman rank correlations in scores and change scores, respectively, between patient-reported outcome measures (sub)scales. The Smallest Detectable Change (SDC) was calculated. The MIC was determined using an anchor-based approach. The presence of floor and ceiling effects was determined.

Results: A total of 140 patients were included. Internal consistency was sufficient for DASH (Cronbach α = 0.96), but was insufficient for Constant-Murley (α = 0.61). Construct and longitudinal validity were sufficient for both patient-reported outcome measures (>75% of correlations hypothesized correctly). The MIC and SDC were 6.7 (95% confidence interval 5.0-15.8) and 19.0 (standard error of measurement, 6.9), respectively, for DASH and 6.1 (95% confidence interval, -6.8 to 17.4) and 17.7 (standard error of measurement, 6.4), respectively, for Constant-Murley.

Conclusions: The DASH and Constant-Murley are valid instruments for evaluating outcome in patients with a humeral shaft fracture. Reliability was only shown for the DASH, making this the preferred instrument. The observed MIC and SDC values provide a basis for sample size calculations for future research.

Level of Evidence: Basic Science Study; Validation of Outcome Instrument

Keywords: Humeral Shaft Fracture; DASH; Constant-Murley; patient reported outcome measure; measurement properties; responsiveness; reliability; and validity

INTRODUCTION

Patient-reported outcome measures (PROMs) are becoming increasingly important instruments to evaluate clinical outcome and functional recovery from the patient's perspective.[1] An advantage of generic quality of life PROMs, such as like the Short Form 36 (SF-36) and EuroQoL-5D (EQ-5D), is that they allow comparison across populations with different medical conditions. Region-specific instruments give insight in disabilities, pain, and problems caused by a specific disease or condition. Some instruments combine a patient-reported part with a clinician-reported part. Effects of treatment can be monitored over time with all three types of instruments, and they can be used to compare different treatment strategies. Instruments should only be used if proven reliable and valid.

The Disabilities of the Arm, Shoulder and Hand (DASH) questionnaire is a region-specific PROM developed in 1996 by a collaborative effort of researchers of the American Academy of Orthopedic Surgeons, and the Institute for Work and Health.[2] It was designed to describe disability experienced by patients with any musculoskeletal condition of the upper extremity and to monitor change in symptoms and upper limb function over time.[3] The DASH outcome measure has been validated in more than 15 languages in patients with a number of upper extremity musculoskeletal disorders, including rheumatoid arthritis and shoulder impingement syndrome.[2, 4] Normative data have been established for the American and Norwegian populations.[2, 5] The Dutch version of the DASH (DASH-DLV) has also been validated in patients with a range of upper extremity disorders.[6]

The Constant-Murley score was developed in 1987 and is currently one of the most-used scales for shoulder (dys)function.[7] The Constant-Murley score evaluates shoulder function by including clinician-assessed physical examination findings and patient-reported assessments. It has been validated for different shoulder pathologies[8, 9] but is also widely used for reporting outcome of patients with a humeral shaft fracture.[10-15]

Although the DASH and Constant-Murley scores have been validated for a number of upper extremity disorders, including shoulder disorders, the measurement properties in the specific population of patients with a humeral shaft fracture are unknown. Also, the Minimal Important Change (MIC) for patients with this injury has not been published before. Knowing this value is important because it may be used as an input parameter for calculating sample sizes for future clinical studies.

The aim of this study was therefore to evaluate the measurement properties of the DASH and Constant-Murley scores in patients who sustained a humeral shaft fracture by comparing them with those of general health-related quality of life instruments subscales (*i.e.*, SF-36 and EuroQoL-5D) and pain measured with a visual analog scale (VAS).

MATERIALS AND METHODS

Data of the first 140 consecutive patients included in a multicenter, prospective cohort study comparing operative and nonoperative treatment of adults with a humeral shaft fracture were used. This study is registered at the Netherlands Trial Register (NTR3617). The study protocol for this trial has been published elsewhere.[16] The medical research ethics committees of all hospitals approved this study, and all patients provided signed informed consent.

Study population

Patients aged 18 years or older presenting with a humeral shaft fracture (AO type 12-A or 12-B) to the Emergency Department of one of 32 participating hospitals in the Netherlands were included. Exclusion criteria were pathological, recurrent, or open fractures, concomitant injuries affecting treatment and rehabilitation of the affected arm, treatment with an external fixator, neurovascular injuries requiring immediate surgery (excluding radial nerve palsy), additional traumatic injuries of the affected arm that influenced upper extremity function, impaired upper extremity function before to the injury, retained hardware around the affected humerus, rheumatoid arthritis, any bone disorder possibly impairing bone healing (excluding osteoporosis), problems of ensuring follow-up (*e.g.*, no fixed address or cognitive impairment), or insufficient comprehension of the Dutch language.

Questionnaires and follow-up measurement

Patients were asked to complete the DASH Dutch language version questionnaire (DASH-DVL),[6] the Constant-Murley score [7], the VAS for the level of pain, EQ-5D,[17] and SF-36[18] at two and six weeks and at three, six, and 12 months after initiation of treatment.

The DASH questionnaire was developed to describe disability experienced by patients with any musculoskeletal condition of the upper extremity and to monitor change in symptoms and upper limb function over time.[4] The DASH questionnaire consists of 30 items, scored 1-5. The DASH score is calculated using the formula: ([sum of all item/number of questions answered] - 1) x 25. The overall score ranges from 0 to 100 points. High scores represent higher disability. Patients needed to have completed at least 27 of 30 of the disability/symptom items of the DASH questionnaire to enable calculation of a total DASH score.[19] The DASH questionnaire has two optional four-item modules enabling measurement of symptoms and upper extremity dysfunction in athletes, performing artists, and other workers whose jobs require more advanced physical activity. These optional modules were not used because they did not apply to the current study population.

The Constant-Murley score evaluates shoulder function by including clinician-assessed physical examination findings and patient-reported assessments.[8] The right and left shoulder are evaluated independently by two clinician-reported items assessing range of motion (ROM) and power and two patient-reported items for pain and activities of daily life (ADL). These are summarized in four dimensions (Constant-Murley pain, ADL, ROM, and power) to create a Constant-Murley total score of 0 to 100 points (15 for pain, 20 for ADL, 40 for ROM, and 25 for power), with a higher score representing a better function. The power subscale was set to zero in patients who were unable to reach 90° abduction or who reported pain during the power measurement. Scores were not normalized to age. Detailed calculations of the Constant-Murley (sub)scales are published elsewhere.[7]

The VAS is used to measure a variety of continuum outcomes. In this study, it was used to measure level of pain. Patients were asked to rate level of pain at each follow-up evaluation by putting a mark on a horizontal line, 100 mm in length, with word descriptors at each end (‘no pain’ at 0 mm and ‘worst pain imaginable’ at 100 mm).[20]

The SF-36 is a validated health survey with 36 questions that represent eight health domains that are combined into a Physical Component Summary (PCS) and a Mental Component Summary (MCS). The score ranges from 0 to 100, with higher scores representing higher quality of life. The scores are converted and compared with the norms for the general population of the United States. The SF-36 is the most widely PROM for assessing general health.[18, 21] A validated Dutch version was used.[22]

The EQ-5D is a standardized instrument for measuring health outcome. It consists of two parts: the EQ-5D utility score (US), and the EQ Visual Analog Scale (EQ-VAS). The EQ-5D US ranges from 0 to 1 and the EQ-VAS ranges from 0 to 100. For both scores, a higher score represents a higher quality of life.[23] A validated Dutch version was used.[17]

Statistical analysis

Data were analyzed using the Statistical Package for the Social Sciences (SPSS) version 21 or higher software (IBM Corp., Armonk, NY, USA). Receiver operating characteristic (ROC) and Youden index were analysed using MedCalc 14.10.2 software (MedCalc Software, Ostend, Belgium). As the raw data for individual items were analyzed, missing data were not imputed.

The measurement properties of the DASH and Constant-Murley scores were determined by comparing them with those of the general health-related quality of life instruments subscales on the SF-36 and EQ-5D and pain measured with a VAS.

Reliability was determined by evaluating internal consistency. The data at six months were used because the largest heterogeneity (ranging from substantial limitation to full recovery) in scores were expected at that time. At an earlier moment, most patients were expected to have substantial functional disability, and at a later time a ceiling effect was expected owing to a large proportion of full recovery. Internal consistency is defined as the extent to which items in a (sub)scale are intercorrelated, thus measuring the same concept.[24] The correlation between items on a (sub)scale was evaluated by calculating the Cronbach α for every (sub)scale. Internal consistency was considered sufficient if the value for Cronbach α was between 0.70 and 0.95, provided that the scale is unidimensional. This analysis requires a sample size of 10 per item in the instrument, with a minimum of 100 patients.[24]

Construct validity represents the extent to which scores on a specific questionnaire relate to other measures in a way that is in agreement with prior theoretically derived hypotheses concerning the concepts that are being measured.[24] The six-months data were used. Continuous data were tested for normality using the Shapiro-Wilk test and by inspecting the quantile-quantile plots. Because the continuous variables were not normally distributed, Spearman rank correlations of the DASH with the (sub)scales of the Constant-Murley score, EQ-5D, and SF-36 scores were calculated to assess DASH construct validity. Correlation coefficients above 0.6, between 0.6 and 0.3, and less than 0.3 were considered high, moderate, and low, respectively.[25] A high correlation between the DASH score and Constant-Murley total and subscale scores with all other (sub)scales or items measuring physical health and functioning (*i.e.*, SF-36 Physical Functioning [PF], SF-36 PCS, EQ-5D ADL, and EQ-5D US) was anticipated. In addition, a moderate-to-low correlation was expected between the SF-36 MCS and the (sub)scales of all other PROMs. A moderate correlation of VAS pain with all other (sub)scales was expected. Finally, we hypothesized that the other individual pain measures (*i.e.*, the Constant-Murley pain subscale, the SF-36 Bodily Pain [BP] subscale, and the EQ-5D pain item) would correlate highly with one another. Construct validity was considered sufficient if at least 75% of the results were in accordance with predefined hypotheses in a (sub)sample of at least 50 patients.[24]

Responsiveness refers to the ability of a questionnaire to detect clinically important changes over time.[24] This was evaluated by assessing longitudinal validity, which refers to the extent to which change in one measurement instrument relates to corresponding change in a reference measure.[26] In addition, the effect size (ES) and standardized response mean (SRM) were determined as measures of the magnitude of change over time.

Longitudinal validity was evaluated by testing predefined hypotheses about expected correlations between DASH and Constant-Murley change scores and the change scores of the EQ-5D and SF-36 (sub)scales. Change scores were calculated as the difference in score from the first to the last follow-up of all instruments that were completed (*i.e.*, six weeks to 12 months). Normality was tested according to the Shapiro-Wilk test and by inspecting the quantile-quantile plots. Correlation coefficients above 0.6, between 0.6 and 0.3, and less than 0.3 were considered high, moderate, and low, respectively.[26] Apart from the Constant-Murley total score, SF-36 BP, and SF-36 PCS, none of the continuous variables showed a normal distribution. Therefore, nonparametric Spearman rank correlations were calculated for all variables of interest. A moderate-to-high correlation between the change scores of the DASH score, the Constant-Murley total score, and the change scores of all other (sub)scales or items measuring physical health and functioning (*i.e.*, SF-36 PF, SF-36 PCS, EQ-5D ADL, and EQ US) was anticipated. A moderate-to-high correlation between the individual pain measures (*i.e.*, Constant-Murley pain subscale, SF-36 BP subscale, and EQ-5D pain item) was expected. Longitudinal validity was considered sufficient if at least 75% of the results were in accordance with predefined hypotheses in a (sub)sample of at least 50 patients.[24]

The ES was calculated by dividing the mean change in score between two time points (*i.e.*, score at 12 months minus the score at six weeks) divided by the standard deviation of the first measurement.[27] The SRM was calculated by dividing the mean change in score between two time points (*i.e.*, score at 12 months minus the score at six weeks) divided by the standard deviation of this change.[27] These effect estimates were interpreted according to Cohen: a value of 0.2 to 0.4 is considered a small, 0.5 to 0.7 a moderate, and ≥ 0.8 a large effect.[25] A large ES was expected a priori because patients were expected to have substantial functional limitations at six weeks, whereas large improvement was expected at 12 months for most patients.

Floor and ceiling effects are present if more than 15% of the study population rates the lowest (floor effect) or highest (ceiling effect) possible score on any PROM (sub)scale.[28] This might limit content validity and responsiveness. In the presence of floor and ceiling effects, items might be missing from the upper or lower ends of the scale, reducing content validity. Likewise, patients with the highest or lowest scores cannot be distinguished from one another, indicating limited reliability.[24] Data of all follow-up moments were evaluated separately.

The MIC represents the smallest measurable change in an outcome score that is perceived significant by patients. This was calculated using an anchor-based method. Patients were asked to complete an ‘anchor question’ or ‘transition item’ at six weeks and at three, six, and 12 months evaluating their perception of change in the general condition of the affected upper limb. The question was: “How is your affected upper arm at this point, in comparison to the previous follow-up moment?” The item scored from 1 “much better” through 2 “a little better”, 3 “more or less the same (no change)”, 4 “a little worse” and 5 “much worse”. The anchor or transition item was considered sufficient if a Spearman rank correlation (r) exceeding 0.29 between the anchor and the change score of the PROM could be demonstrated.[29] The change score (score at last follow-up minus the score at completion of the transition item) of patients who selected “a little better” on the transition item was considered the MIC.

The MIC was calculated for the total scores by plotting the ROC curve of the change in score for patients who scored “a little better” on the transition item compared with patients who scored “more or less the same (no change).” The area under the ROC curve is provided as a measure of discriminatory power. The optimal ROC cutoff point calculated with the Youden index reflected the value of the MIC. The Youden index is shown with its 95% confidence interval (CI) after bootstrapping (1,000 replicates and 900 random-number seeds).

The smallest intrapersonal change in score that represents (with $P < .05$) a “real” difference above measurement error is defined by the Smallest Detectable Change (SDC) of a measurement instrument.[1] This was based on the change scores of patients who answered “more or less the same/no change” on the transition item; patients were assumed to be stable in the interim period. For the individual patient, the SDC was derived from the standard error of measurement (SEM) according to the following formula: $SDC = 1.96 \times \sqrt{2} \times SEM$. SEM was calculated as $SD_{change} / \sqrt{2}$. Ideally, for evaluative purposes, the SDC should be smaller than the MIC[24]

RESULTS

Study population

This study population comprised 140 patients who sustained a humeral shaft fracture; of these, 19 patients were lost to follow-up (four after two weeks follow-up, five after six weeks, six after three months, and four after six months). In addition, seven patients missed one follow-up visit (five missed two weeks, one six weeks, and one six months). The median age was 58 years (25th percentile-75th percentile, 41-68) and 63 patients (45.0%) were male. The right arm was affected in 65 patients (46.4%), and the dominant arm was affected in 64 patients (45.7%).

The changes over time in DASH, Constant-Murley total and subscales, and VAS pain of patients with a humeral shaft fracture are shown in Fig. 1. All scores showed a decrease in symptoms, disability,y or pain over time, except for the Constant-Murley pain subscale, which displayed a similar score at all follow-up assessments. The change in SF-36 PCS, SF-36 MCS, EQ-5D US, and EQ-5D VAS scores over time is shown in Fig 2. The PROM (sub)scales scores measuring physical health and general health (SF-36 PCS and EQ-5D US) increased over time, but the mental health-related quality of life and the perception of health-related quality of life state (SF-36 MCS and EQ-5D VAS) were stable over time.

Reliability

The Cronbach α value of DASH score (α = 0.96) was sufficient, indicating high correlation among the 30 items (Table 1). Cronbach α values of the Constant-Murley ROM subscale (α=0.88) also indicated sufficient internal consistency. Internal consistency of the Constant-Murley total score (α = 0.61) and the Constant-Murley ADL subscale (α = 0.60) was

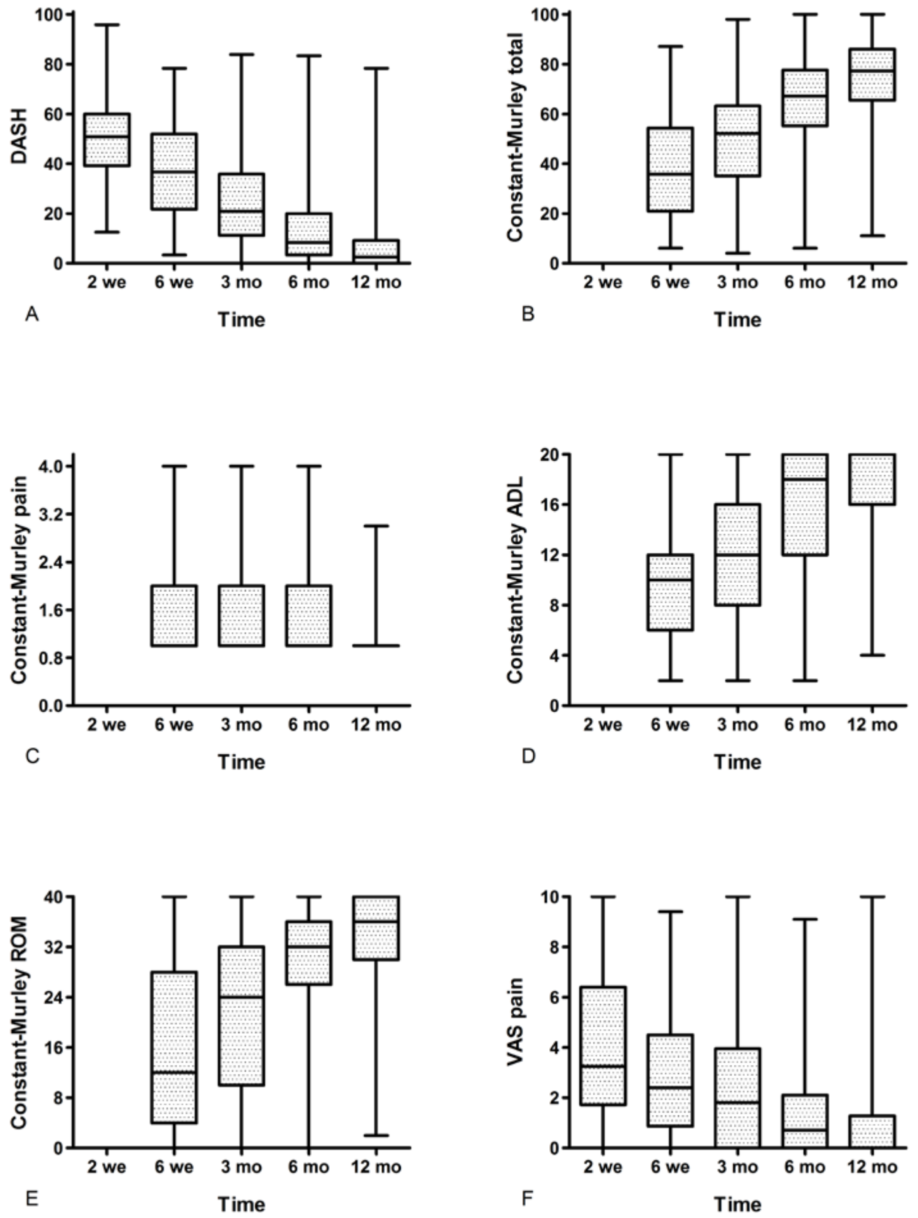

Figure 1 (A) Disabilities of the Arm, Shoulder and Hand (DASH), (**B**) Constant-Murley total, (**C**) Constant-Murley pain, (**D**) Constant-Murley Activities of Daily Life (ADL), (**E**) Constant-Murley Range of Motion (ROM), and(**F**) visual analog scale(VAS) pain scores at each follow-up visit in patients with a humeral shaft fracture. The horizontal line in the middle of each box indicates the median, the top and bottom borders of the box mark the 75th and 25th percentiles, respectively, and the whiskers mark the 90th and 10th percentiles.

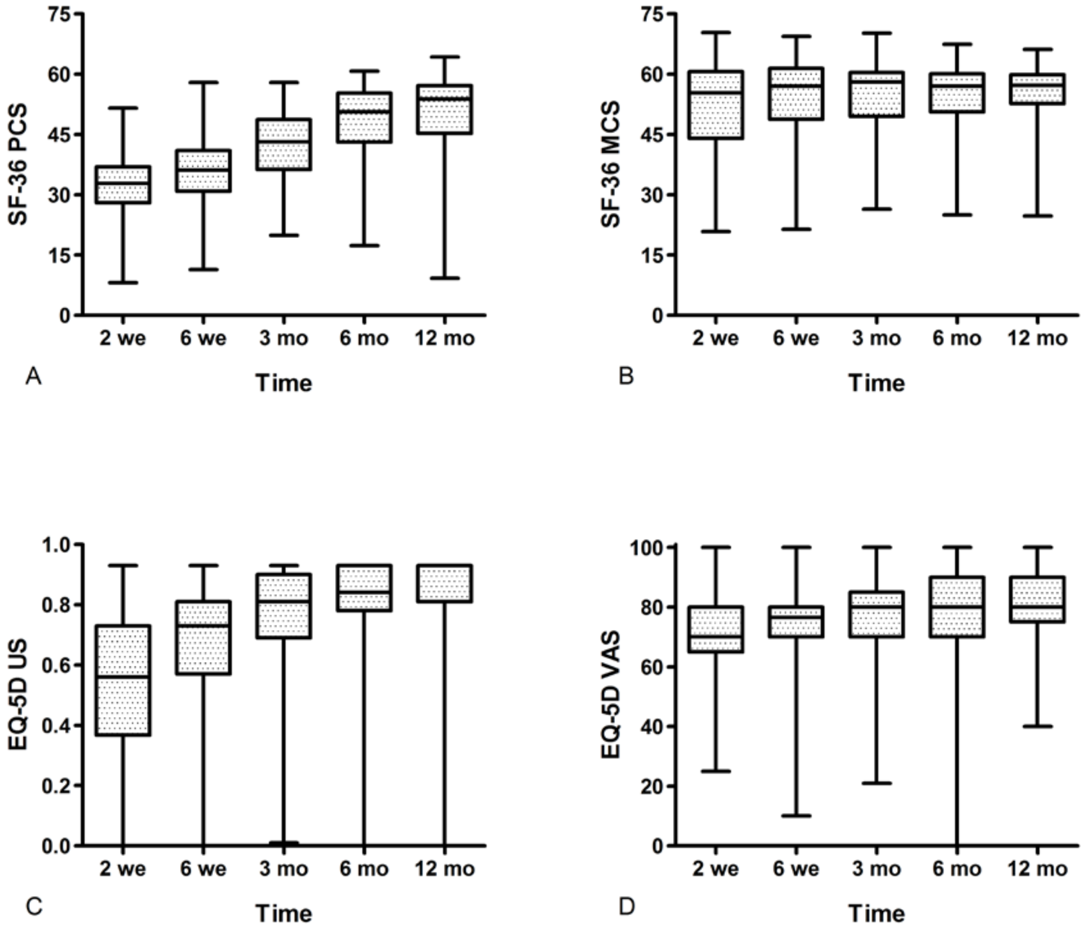

Figure 2 (**A**) Short Form 36 (SF-36) Physical Component Summary (PCS), (**B**) SF-36 Mental Component Summary MCS, (**C**) EuroQoL-5D (EQ-5D) Utility Score (US), and (**D**) EQ-5D visual analog scale (VAS) scores at each follow-up visit in patients with a humeral shaft fracture. The horizontal line in the middle of each box indicates the median, the top and bottom borders of the box mark the 75th and 25th percentiles, respectively, and the whiskers mark the 90th and 10th percentiles.

Table 1 Internal consistency of the instruments in patients with a humeral shaft fracture*

Instrument	No.	No. of items	Cronbach α
DASH (all items)	115	30	0.96
Constant-Murley (all items)	115	10	0.61†
ADL	122	4	0.60
ROM	122	4	0.88
Pain	122	1	N.D. ‡
Power	115	1	N.D. ‡

ADL, activities of daily life; DASH, Disabilities of the Arm, Shoulder and Hand; N.D., not determined; ROM, range of motion.

* Data are shown for the six months' follow-up. The maximum number of patients was 125.

† Value should be interpreted carefully because the total scale is not unidimensional.

‡ The Constant-Murley pain and power subscales consist of single items. Internal consistency does not apply to a single-item domain.

insufficient. No Cronbach α was determined for the Constant-Murley pain and power subscales, because internal consistency does not apply to a single-item domain.

Construct validity

Construct validity is presented in Table 2. The calculated Spearman rank correlations confirmed 12 of 14 prior hypothesized correlations (85.7%) between the DASH and (sub)scales of the other PROMs, indicating sufficient construct validity. The construct validity was sufficient for the Constant-Murley total score, and Constant-Murley power (11 of 14 [78.6%]) was also sufficient. However, construct validity for the other subscales was not sufficient. A high correlation of the DASH score and the Constant-Murley total score was found with the subscales of other PROMs focusing on physical health and functioning (*i.e.*, SF-36 PCS, SF-36 PF, and EQ-5D US). The DASH showed a moderate correlation with the SF-36 MCS, whereas the Constant-Murley total and subscale scores showed low correlations with SF-36 MCS. The moderate correlation between the Constant-Murley pain subscale and the VAS pain score was hypothesized correctly, but the moderate correlation with the other individual pain measures (*i.e.*, SF-36 BP subscale and EQ-5D pain item) contradicted the predefined hypotheses. The moderate correlation between the Constant-Murley ROM subscale and EQ-5D ADL scores was also not expected.

Table 2 Construct validity of the instruments in patients with a humeral shaft fracture*

Variable	DASH	Constant-Murley				
		Total	Pain	ADL	ROM	Power
DASH	1	-0.78 [114]	0.52 [121]	-0.71 [121]	-0.60 [121]	-0.57 [114]
Constant-Murley						
(total score)	-0.78 [114]	1	-0.52 [115]	0.72 [115]	0.89 [115]	0.82 [115]
Pain	0.52 [121]	-0.52 [115]	1	-0.45 [122]	-0.31 [122]	-0.24 [115]
ADL	-0.71 [121]	0.72 [115]	-0.45 [122]	1	**0.48 [122]**	0.45 [115]
ROM	-0.60 [121]	0.89 [115]	-0.31 [122]	**0.48 [122]**	1	**0.69 [115]**
Power	-0.57 [114]	0.82 [115]	-0.24 [115]	0.45 [115]	0.69 [115]	1
VAS Pain	0.72 [123]	-0.53 [115]	0.57 [122]	-0.48 [122]	-0.34 [122]	**-0.40 [115]**
SF-36 PCS	-0.79 [121]	0.65 [112]	-0.38 [119]	**0.55 [119]**	**0.50 [119]**	0.54 [112]
SF-36 MCS	-0.31 [121]	0.14 [112]	-0.14 [119]	0.11 [119]	0.03 [119]	0.09 [112]
PF	-0.73 [123]	0.65 [114]	**-0.27 [121]**	**0.46 [121]**	**0.53 [121]**	0.59 [114]
BP	-0.65 [123]	**0.46 [114]**	**-0.55 [121]**	0.40 [121]	**0.27 [121]**	0.38 [114]
EQ-5D US	-0.67 [123]	0.60 [114]	-0.33 [121]	**0.43 [121]**	0.42 [121]	0.55 [114]
ADL	-0.60 [123]	**0.53 [114]**	-0.30 [121]	**0.55 [121]**	**0.38 [121]**	0.35 [114]
Pain	**-0.57 [123]**	0.44 [114]	**-0.40 [121]**	0.35 [121]	**0.24 [121]**	**0.38 [114]**
VAS	**-0.53 [123]**	**0.48 [114]**	**-0.24 [121]**	**0.31 [121]**	0.37 [121]	0.45 [114]

ADL, activities of daily life; BP, bodily pain; DASH, Disabilities of the Arm, Shoulder and Hand; EQ-5D, EuroQoL-5D; MCS, Mental Component Summary; PCS, Physical Component Summary; PF, Physical Functioning; ROM, range of motion; SF-36, Short Form-36; VAS, Visual Analog Scale; US, utility score.

* Data are shown for the six months' follow-up. The maximum number of patients was 125. Construct validity is shown as Spearman rank correlation coefficients (r) with brackets showing the number of patients included in the correlation: r > 0.6 high correlation, r = 0.3 to 0.6 moderate correlation, and r < 0.3 low correlation. Bold correlations were not hypothesized correctly.

Responsiveness

Longitudinal validity is presented in Table 3. The DASH score demonstrated sufficient longitudinal validity, with 11 of 14 change score correlations (78.6%) hypothesized correctly. As anticipated, a high correlation was found between the change scores of the DASH, the Constant-Murley total, and Constant-Murley ADL subscale scores. The moderate correlation between the DASH and the SF-36 PCS and SF-36 PF was not expected. The low correlation between the DASH and the Constant-Murley power subscale was also not expected.

The longitudinal validity of the Constant-Murley total score was sufficient. Of the 14 hypotheses, (85.7%) 12 were correct. The high correlation with the DASH and Constant-Murley ADL and ROM subscales was as expected. The moderate correlation with the SF-36 PCS and PF was not expected. The individual Constant-Murley subscales of pain, ADL, ROM, and power showed insufficient longitudinal validity, with 57.1%, 71.4%, 64.3% and 64.3% correct hypotheses, respectively.

The SRM and the ES of the DASH and Constant-Murley instruments are reported in Table 4. The magnitude of change over time was large for the DASH and Constant-Murley total and ADL, ROM, and power subscales (SRM and ES >1.3). The magnitude of change for the Constant-Murley pain subscale was medium (SRM -0.58 and ES -0.64).

Table 3 Longitudinal validity of the instruments in patients with a humeral shaft fracture*

	DASH	Constant-Murley				
		Total	Pain	ADL	ROM	Power
DASH	1	-0.60 [104]	0.45 [114]	-0.64 [114]	-0.54 [114]	**-0.14 [104]**
Constant-Murley						
Total score	-0.60 [104]	1	-0.43 [105]	0.76 [105]	0.90 [105]	0.53 [105]
Pain	0.45 [114]	-0.43 [105]	1	**-0.26 [116]**	**-0.29 [115]**	-0.12 [105]
ADL	-0.64 [114]	0.76 [105]	**-0.26 [116]**	1	0.70 [115]	0.23 [105]
ROM	-0.54 [114]	0.90 [105]	**-0.29 [115]**	0.70 [115]	1	**0.29 [105]**
Power	**-0.14 [104]**	0.53 [105]	-0.12 [105]	0.23 [105]	**0.29 [105]**	1
VAS Pain	0.55 [118]	-0.46 [105]	**0.45 [116]**	-0.46 [116]	-0.33 [115]	-0.18 [105]
SF-36 PCS	**-0.56 [116]**	**0.54 [102]**	-0.40 [112]	**0.52 [112]**	**0.48 [112]**	**0.24 [102]**
SF-36 MCS	-0.20 [116]	0.02 [102]	0.01 [112]	0.01 [112]	0.02 [112]	-0.07 [102]
PF	**-0.57 [117]**	**0.34 [103]**	**-0.16 [113]**	**0.34 [113]**	**0.40 [113]**	**0.07 [103]**
BP	-0.47 [118]	0.40 [104]	**-0.36 [115]**	0.42 [115]	0.37 [114]	0.12 [104]
EQ-5D US	-0.55 [118]	0.51 [104]	-0.25 [115]	0.40 [115]	0.46 [114]	0.09 [104]
ADL	-0.50 [118]	0.44 [104]	-0.19 [115]	**0.41 [115]**	0.34 [114]	**0.21 [104]**
Pain	-0.41 [118]	0.35 [104]	**-0.43 [115]**	0.34 [115]	**0.26 [114]**	0.18 [104]
VAS	-0.18 [118]	0.25 [104]	-0.23 [115]	0.18 [115]	0.15 [114]	0.15 [104]

ADL, activities of daily life; BP, bodily pain; DASH, Disabilities of the Arm, Shoulder and Hand; EQ-5D, EuroQoL-5D; MCS, Mental Component Summary; PCS, Physical Component Summary; PF, Physical Functioning; ROM, range of motion; SF-36, Short Form-36; VAS, Visual Analog Scale; US, utility score.

* Responsiveness is shown as Spearman rank correlation coefficients (r) of change in scores between six weeks and 12 months with the number of patients included in the correlation between brackets. The maximum number of patients was 121. Values of $r > 0.6$ indicate high correlation, $r = 0.3$ to 0.6 indicate moderate correlation, and $r < 0.3$ indicate low correlation. The bold correlations were not hypothesized correctly.

Table 4 Responsiveness: standardized response mean and effect size of the instruments in patients with a humeral shaft fracture*

Instrument	No.	Mean change	SD_{change}	SRM	$SD_{6\ weeks}$	ES
DASH	118	-27.8	17.1	-1.63	18.0	-1.55
Constant-Murley	105	34.2	21.4	1.60	20.0	1.71
Pain	116	-0.5	0.9	-0.58	0.8	-0.64
ADL	116	8.6	4.8	1.78	4.3	2.01
ROM	115	17.7	13.0	1.36	13.1	1.35
Power	105	6.9	6.3	1.10	4.0	1.75

ADL, activities of daily life; DASH, Disabilities of the Arm, Shoulder and Hand; ES, effect size; ROM, range of motion; SD, standard deviation of mean change; SRM, standardized response mean.

* Change scores were calculated from six weeks to 12 months. The maximum number of patients was 121.

Floor and ceiling effects

Floor effects were not present in the DASH and Constant-Murley total and ADL and ROM subscale scores at any of the follow-up assessments (Fig. 3, A). However, floor effects were present in the Constant-Murley pain subscale at all follow-up assessments.

A ceiling effect was seen for the DASH score at 12 months of follow-up, with 31.1% of patients reporting no disability at that assessment (Fig. 3, B). For the Constant-Murley ADL and ROM subscale scores, ceiling effects were demonstrated at six and 12 months.

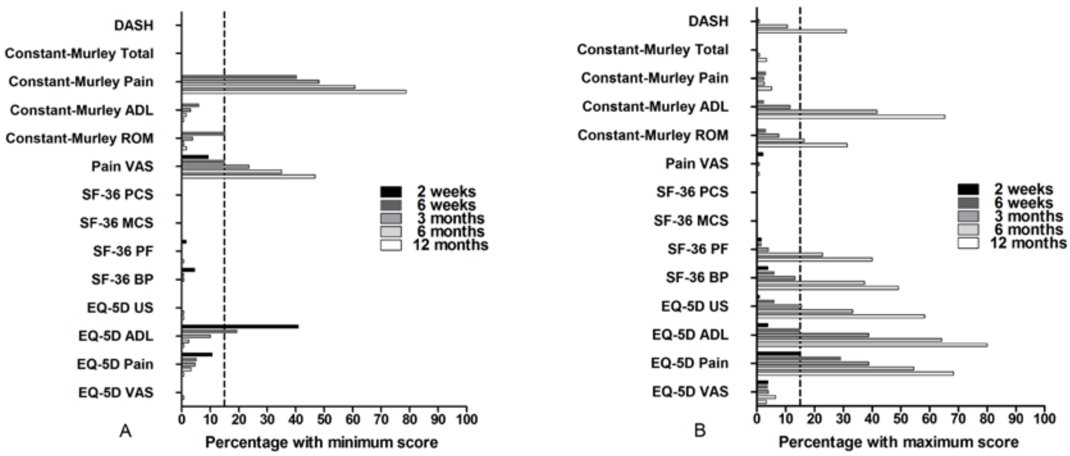

Figure 3 (A) Floor and(B) ceiling effects of the instruments at each follow-up visit in patients with a humeral shaft fracture. ADL, activities of daily life; BP, bodily pain; DASH, Disabilities of the Arm, Shoulder and Hand; EQ-5D, EuroQoL-5D; ROM, range of motion; SF-36, Short Form-36; MCS, Mental Component Summary; PCS, Physical Component Summary; PF, Physical Functioning; US, utility score; VAS, visual analog scale.

MIC and SDC

Anchor-based MIC and distribution-based SDC values are given in Table V. Thirty percent of transition items were reported as “a little better” and 14.4% as “more or less the same (no change).” The transition item displayed a sufficient correlation (*i.e.*, r > 0.3) with the change scores of the DASH, Constant-Murley total scores, as well as with the Constant-Murley ADL and ROM subscales. Insufficient Spearman rank correlations with the transition item was found for the change scores of the Constant-Murley pain subscale (r = 0.21) and power subscale (r = -0.18); therefore the MIC for the pain and power subscale could not be determined. The MIC value was 6.7 (95% CI, 5.0-15.8) for the DASH score and 6.1 (95% CI -6.8 to 17.4) for the Constant-Murley score. The MIC was smaller than the SDC for all total and subscale scores. The SDC was 19.0 (SEM, 6.9) for the DASH score and 17.7 (SEM 6.4) for the Constant-Murley score.

Table 5 Minimal Important Change and Smallest Detectable Change values of the instruments in patients with a humeral shaft fracture

Instrument	Scoring range	Anchor-based approach* No.	AUC	MIC	Sensitivity (%)	Specificity (%)	Distribution-based approach* No.	SD_{change}	SEM	SDC
DASH	0-100	150	0.66 (0.58-0.73)	6.7 (5.0-15.8)	45.3	80.8	73	9.7	6.9	19.0
Constant-Murley (total score)	0-100	105	0.59 (0.50-0.68)	6.1 (-6.8 to 17.4)	58.1	61.8	55	9.0	6.4	17.7
Pain	0-15	120	0.52 (0.44-0.59)	N.D.	N.D.	N.D.	59	0.6	0.4	1.2
ADL	0-20	120	0.59 (0.51-0.67)	N.D.	N.D.	N.D.	58	2.4	1.7	4.7
ROM	0-40	120	0.62 (0.54-0.70)	N.D.	N.D.	N.D.	58	5.1	3.6	9.9
Power	0-25	105	0.57 (0.48-0.66)	N.D.	N.D.	N.D.	55	4.1	2.9	8.0

ADL, activities of daily life; AUC, area under the curve; DASH, Disabilities of the Arm, Shoulder, and Hand; MIC, minimal important change; N.D., not determined; ROM, range of motion; SD, standard deviation of mean change; SDC, smallest detectable change; SEM, standard error of measurement.

* Anchor-based and distribution-based methods for MIC and SDC values, respectively. For the MIC, the area under the receiver operating characteristic curve and MIC are shown with 95% confidence intervals between brackets.

DISCUSSION

Results of the current study show that the DASH and Constant-Murley are valid instruments to describe symptoms and disability experienced by patients who sustained a humeral shaft fracture over time. The DASH was also found to be reliable.

The DASH instrument and the Constant-Murley ROM subscale demonstrated sufficient internal consistency in this population, as reflected by Cronbach α values of at least 0.70. The observed value for the DASH was consistent with previously published values, which range from 0.91 to 0.98.[30, 31] The Cronbach α, however, exceeded 0.95, suggesting that some of the items of the DASH questionnaire might be redundant for adequate construct measurement in this research setting. The internal consistency of the Constant-Murley total score of 0.61 was within the range of 0.60 to 0.75 described previously.[8] The value should be interpreted carefully because the total instrument is multidimensional. The insufficient internal consistency of the Constant-Murley ADL subscale was a novel finding. However, because the Cronbach α is dependent on the number of items in a (sub)scale, the inferior result might be related to the small number of items (three items) in the Constant-Murley ADL subscale.[24]

Construct validity of the DASH score was sufficient, with 85.7% of the predicted correlations confirmed. More specifically, the DASH displayed high correlations with the Constant-Murley total score, the Constant-Murley ADL and ROM subscales, and subscales of other PROMs focusing on physical health and functioning. The unexpected low correlation between the DASH and the Constant-Murley power subscale may suggest that not all activities asked in the DASH are affected by differences in power of the shoulder. The high correlation between the DASH and the EQ-5D has been published in patients with a proximal humeral fracture and was of comparable strength.[3] To the contrary, the correlations between the DASH and the SF-36 MCS found in this study was much lower than previously described.[32] The unexpected moderate correlation between the DASH and the SF-36 PCS and PF may be because patients more often than expected had functional limitations caused by conditions not affecting the upper extremity; these affect the SF-36 but not the DASH. Interestingly, only a moderate correlation was found between the DASH and the EQ-5D VAS. This suggests that sustaining a humeral shaft fracture does not necessarily affect a patient's general health perception. Cederlund *et al.* reported a similar finding in patients who received treatment for major hand surgery. The patients in their study had the same median general health perception as scored by the EQ-5D VAS at three and six months after initiation of treatment.[33]

According to Cohen's[25] interpretation, the SRM values of the DASH (-1.63), the Constant-Murley total score (1.60) and its (sub)scale scores suggested good to excellent ability to detect clinical change over time.[25] Other studies reported SRM values for the DASH in different contexts, with values ranging from -0.48 to -1.64.[34, 35] No published SRM values for the Constant-Murley score were found.

In this study, the DASH and Constant-Murley scores displayed sufficient longitudinal validity as reflected by 78.6% and 85.7% of correctly hypothesized correlations, respectively. Correlations in change scores of the DASH with the SF-36 PCS and SF-36 MCS were comparable to a previous study. [36]

The DASH score displayed a ceiling effect at 12 months' follow-up. Treatment of humeral shaft fractures is aimed at full recovery, and achieving this will cause a ceiling effect because patients who have a full recovery report no disabilities on PROMs. In this study, population full recovery of a substantial portion of the patients was expected one year after the start of treatment, and so a ceiling effect was expected. But because of the ceiling effect, differences in the group of patients who reported no disabilities at 12 months' follow-up cannot be distinguished, making it not suitable to, for example, use this time point to compare differences of different treatment strategies.

The anchor-based MIC for the DASH was 6.7 (95% CI, 5.0-15.8), which is a little lower than found in previous studies. Previously published MIC values range from seven in patients who sustained ulnar nerve decompression to 15 in patients with shoulder impingement syndrome.[33, 37] Because MIC values are known to be patient and context dependent, it is likely that the differences in study populations explain the differences in reported MIC values.[24] MICfor the Constant-Murley score has not been reported previously.[8] The SDC as found for this instrument in the current study, 17.7 points, is in line with 17 to 23 points reported previously for shoulder impingement, supraspinatus tears, and massive rotator cuff tears._ENREF_13[38] For monitoring changes in individual patients (*e.g.* in clinical practice), the MIC should be larger than the SDC. This is necessary to make a distinction between "real' change and change induced by measurement error. In research, however, the MIC is used differently (*e.g.*, to determine percentages of responders)', and the measurement error is much smaller. For all PROM (sub)scales in this study, the anchor-based MIC was smaller than the SDC. This suggests that the observed MIC values in this study fall into the range that could be due to chance.

This study has some limitations. Because there was too much time between two subsequent follow-up moments, performing an adequate test-retest analysis was not possible. Therefore, calculation of the SEM was done with the corresponding change scores of patients who answered "no change" on the transition item. This may have resulted in incorrect SEM, because the Spearman rank correlations between the transition item and change scores of the Constant-Murley pain subscale was insufficient. For the other items, however, the correlation was sufficient, so this did not apply to those items. Similarly, this may have hindered correct anchor-based MIC and SDC calculations. As a second limitation, the calculations were done using the non-normalized Constant-Murley scores because the sample size did not allow stratification by age.

CONCLUSIONS

This study confirms, for the first time, that the DASH and Constant-Murley scores are valid for evaluating outcome over time in patients who sustained a humeral shaft fracture. Reliability was confirmed only for the DASH, making this the most suitable instrument. Ceiling effects were noted at one-year follow-up, likely owing to increasing numbers of patients with full recovery. For the DASH, the MIC was 6.7 (95% CI, 5.0-15.8) and the SDC was 19.0 (SEM, 6.9). For the Constant-Murley score, the MIC was 6.1 (95% CI, -6.8 to 17.4) and the SDC was 17.7 (SEM, 6.4). The MIC and SDC values enable adequate sample size calculations for future research.

REFERENCES

1. Davidson M, Keating J. Patient-reported outcome measures (PROMs): how should I interpret reports of measurement properties? A practical guide for clinicians and researchers who are not biostatisticians. British Journal of Sports Medicine. 2014 May;48(9):792-6. Epub 2012/12/22.

2. Wylie JD, Beckmann JT, Granger E, Tashjian RZ. Functional outcomes assessment in shoulder surgery. World Journal of Orthopedics. 2014 Nov 18;5(5):623-33. Epub 2014/11/19.

3. Slobogean GP, Noonan VK, O'Brien PJ. The reliability and validity of the Disabilities of Arm, Shoulder, and Hand, EuroQol-5D, Health Utilities Index, and Short Form-6D outcome instruments in patients with proximal humeral fractures. J Shoulder Elb Surg. 2010 Apr;19(3):342-8. Epub 2010/03/02.

4. Hudak PL, Amadio PC, Bombardier C. Development of an upper extremity outcome measure: the DASH (disabilities of the arm, shoulder and hand) [corrected]. The Upper Extremity Collaborative Group (UECG). American Journal of Industrial Medicine. 1996 Jun;29(6):602-8. Epub 1996/06/01.

5. 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 Musculoskel Dis. 2014 Mar 12;15:78. Epub 2014/03/13.

6. Veehof MM, Sleegers EJ, van Veldhoven NH, Schuurman AH, van Meeteren NL. Psychometric qualities of the Dutch language version of the Disabilities of the Arm, Shoulder, and Hand questionnaire (DASH-DLV). Journal of Hand Therapy. 2002 Oct-Dec;15(4):347-54. Epub 2002/11/27.

7. Constant CR, Murley AHG. A Clinical Method of Functional Assessment of the Shoulder. Clin Orthop Relat R. 1987 Jan;214:160-4. Epub 1987/01/01.

8. Roy JS, MacDermid JC, Woodhouse LJ. A systematic review of the psychometric properties of the Constant-Murley score. J Shoulder Elb Surg. 2010 Jan;19(1):157-64. Epub 2009/06/30.

9. Blonna D, Scelsi M, Marini E, Bellato E, Tellini A, Rossi R, et al. Can we improve the reliability of the Constant-Murley score? J Shoulder Elb Surg. 2012 Jan;21(1):4-12. Epub 2011/10/19.

10. Brunner A, Thormann S, Babst R. Minimally invasive percutaneous plating of proximal humeral shaft fractures with the Proximal Humerus Internal Locking System (PHILOS). J Shoulder Elbow Surg. 2012 Aug;21(8):1056-63. Epub 2011/08/30.

11. Fan Y, Li YW, Zhang HB, Liu JF, Han XM, Chang X, et al. Management of Humeral Shaft Fractures With Intramedullary Interlocking Nail Versus Locking Compression Plate. Orthopedics. 2015 Sep;38(9):e825-9. Epub 2015/09/17.

12. Li Y, Wang C, Wang M, Huang L, Huang Q. Postoperative malrotation of humeral shaft fracture after plating compared with intramedullary nailing. J Shoulder Elbow Surg. 2011 Sep;20(6):947-54. Epub 2011/03/29.

13. Muckley T, Diefenbeck M, Sorkin AT, Beimel C, Goebel M, Buhren V. Results of the T2 humeral nailing system with special focus on compression interlocking. Injury. 2008 Mar;39(3):299-305. Epub 2008/01/01.

14. Tsourvakas S, Alexandropoulos C, Papachristos I, Tsakoumis G, Ameridis N. Treatment of humeral shaft fractures with antegrade intramedullary locking nail. Musculoskelet Surg. 2011 Dec;95(3):193-8. Epub 2011/07/12.

15. Wang C, Li J, Li Y, Dai G, Wang M. Is minimally invasive plating osteosynthesis for humeral shaft fracture advantageous compared with the conventional open technique? J Shoulder Elbow Surg. 2015 Nov;24(11):1741-8. Epub 2015/10/21.

16. Mahabier KC, van Lieshout EMM, Bolhuis HW, Bos PK, Bronkhorst MWGA, Bruijninckx MMM, et al. HUMeral Shaft Fractures: MEasuring Recovery after Operative versus Non-operative Treatment (HUMMER): a multicenter comparative observational study. Bmc Musculoskel Dis. 2014 Feb 11;15:39. Epub 2014/02/13.

17. EuroQol G. EuroQol--a new facility for the measurement of health-related quality of life. Health Policy. 1990 Dec;16(3):199-208. Epub 1990/11/05.

18. Ware JE, Jr., Sherbourne CD. The MOS 36-item short-form health survey (SF-36). I. Conceptual framework and item selection. Medical Care. 1992 Jun;30(6):473-83. Epub 1992/06/11.

19. Kennedy CA, Beaton DE, Solway S, McConnell S, Bombardier C. Disabilities of the Arm, Shoulder and Hand (DASH). The DASH and QuickDASH Outcome Measure User’s Manual. Toronto, Ontario: Institute for Work & Health; 2011.

20. Wewers ME, Lowe NK. A Critical-Review of Visual Analog Scales in the Measurement of Clinical Phenomena. Research in Nursing & Health. 1990 Aug;13(4):227-36. Epub 1990/08/01.

21. Garratt AN, Schmidt L, Mackintosh A, Fitzpatrick R. Quality of life measurement: bibliographic study of patient assessed health outcome measures. Brit Med J. 2002 Jun 15;324(7351):1417-9. Epub 2002/06/18.

22. Aaronson NK, Muller M, Cohen PDA, Essink-Bot ML, Fekkes M, Sanderman R, et al. Translation, validation, and norming of the Dutch language version of the SF-36 Health Survey in community and chronic disease populations. Journal of Clinical Epidemiology. 1998 Nov;51(11):1055-68. Epub 1998/11/17.

23. Norman R, Cronin P, Viney R, King M, Street D, Ratcliffe J. International Comparisons in Valuing EQ-5D Health States: A Review and Analysis. Value in Health. 2009 Nov-Dec;12(8):1194-200. Epub 2009/08/22.

24. 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. Journal of Clinical Epidemiology. 2007 Jan;60(1):34-42. Epub 2006/12/13.

25. Cohen J. Statistical power analysis for the behavioral sciences. New York: Academic Press; 1977. p. 474.

26. Middel B, van Sonderen E. Statistical significant change versus relevant or important change in (quasi) experimental design: some conceptual and methodological problems in estimating magnitude of intervention-related change in health services research. International Journal of Integrated Care. 2002;2:e15. Epub 2006/08/10.

27. Angst F, Verra ML, Lehmann S, Aeschlimann A. Responsiveness of five condition-specific and generic outcome assessment instruments for chronic pain. BMC Med Res Methodol. 2008;8:26. Epub 2008/04/29.

28. McHorney CA, Tarlov AR. Individual-patient monitoring in clinical practice: are available health status surveys adequate? Quality of Life Research. 1995 Aug;4(4):293-307. Epub 1995/08/01.

29. Revicki D, Hays RD, Cella D, Sloan J. Recommended methods for determining responsiveness and minimally important differences for patient-reported outcomes. Journal of Clinical Epidemiology. 2008 Feb;61(2):102-9. Epub 2008/01/08.

30. Kitis A, Celik E, Aslan UB, Zencir M. DASH questionnaire for the analysis of musculoskeletal symptoms in industry workers: A validity and reliability study. Applied Ergonomics. 2009 Mar;40(2):251-5. Epub 2008/06/17.

31. Dias JJ, Rajan RA, Thompson JR. Which questionnaire is best? The reliability, validity and ease of use of the Patient Evaluation Measure, the Disabilities of the Arm, Shoulder and Hand and the Michigan Hand Outcome Measure. J Hand Surg-Eur Vol. 2008 Feb;33(1):9-17. Epub 2008/03/12.

32. SooHoo NF, McDonald AP, Seiler JG, McGillivary GR. Evaluation of the construct validity of the DASH questionnaire by correlation to the SF-36. J Hand Surg-Am. 2002 May;27A(3):537-41. Epub 2002/05/17.

33. Cederlund RI, Ramel E, Rosberg HE, Dahlin LB. Outcome and clinical changes in patients 3, 6, 12 months after a severe or major hand injury - can sense of coherence be an indicator for rehabilitation focus? Bmc Musculoskel Dis. 2010 Dec 22;11:286. Epub 2010/12/24.

34. Ishikawa H, Murasawa A, Nakazono K, Abe A, Otani H, Netsu T, et al. The patient-based outcome of upper-extremity surgeries using the DASH questionnaire and the effect of disease activity of the patients with rheumatoid arthritis. Clinical Rheumatology. 2008 Aug;27(8):967-73. Epub 2008/01/25.

35. Aasheim T, Finsen V. The DASH and the QuickDASH instruments. Normative values in the general population in Norway. J Hand Surg-Eur Vol. 2014 Feb;39(2):140-4. Epub 2013/03/23.

36. MacDermid JC, Drosdowech D, Faber K. Responsiveness of self-report scales in patients recovering from rotator cuff surgery. J Shoulder Elb Surg. 2006 Jul-Aug;15(4):407-14. Epub 2006/07/13.

37. Lee JY, Lim JY, Oh JH, Ko YM. Cross-cultural adaptation and clinical evaluation of a Korean version of the disabilities of arm, shoulder, and hand outcome questionnaire (K-DASH). J Shoulder Elb Surg. 2008 Jul-Aug;17(4):570-4. Epub 2008/05/13.

38. Henseler JF, Kolk A, van der Zwaal P, Nagels J, Vliet Vlieland TP, Nelissen RG. The minimal detectable change of the Constant score in impingement, full-thickness tears, and massive rotator cuff tears. J Shoulder Elbow Surg. 2015 Mar;24(3):376-81. Epub 2014/09/23.