

Psychometric properties of two questionnaires in the context of total wrist arthroplasty

Michel E.H. Boeckstyns¹ & Søren Merser²

ABSTRACT

INTRODUCTION: Patient-rated outcome measures are frequently used to assess the results of total wrist arthroplasty, but their psychometric properties have not yet been evaluated in this group of patients. The purpose of our study was to assess the psychometric properties of the Danish Quick Disabilities of Arm Shoulder and Hand (QuickDASH) and Patient-rated Wrist Evaluation questionnaires in patients with total wrist arthroplasty.

MATERIAL AND METHODS: In a prospective cohort of 102 cases, we evaluated the QuickDASH. Furthermore, in a cross-sectional study and a test-retest on a subgroup of the patients, we evaluated the Patient-rated Wrist Evaluation.

RESULTS: Internal consistency and reproducibility were very high (Cronbach's alpha 0.96/0.97; Spearman's rho 0.90/0.91; intraclass coefficient 0.91/0.92), and there were no floor or ceiling effects. The responsiveness of the QuickDASH was high (standardised response mean 1.06 and effect size 1.07). The construct validity of both scales was confirmed by three a priori formulated hypotheses: a moderate, negative correlation of scores with grip-strength; a moderate, positive correlation with pain and a very weak or no correlation with mobility. Rheumatoid patients scored significantly higher on the QuickDASH than other patients did. The scores of both questionnaires were very closely related.

CONCLUSION: Both questionnaires are valid and equivalent for use in patients with total wrist arthroplasty.

FUNDING: This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

TRIAL REGISTRATION: not relevant.

The QuickDASH questionnaire was developed by extracting 11 of 30 items from the original DASH questionnaire. It aims at measuring function, disability and symptoms in persons with disorders of the upper limb with a short patient-rated outcome instrument [1]. In a systematic review, Kennedy et al identified studies validating the original English version and cultural adaptations [2]. The diagnostic groups in these studies vary widely. In one study only, patients with upper limb arthroplasties were included, but none of them had wrist arthroplasties [3]. The Patient-rated Wrist Evaluation questionnaire (PRWE) [4] was originally designed as a specific instru-

ment for the assessment of distal radius fractures and wrist injuries, but this questionnaire has not been validated in the specific context of wrist arthroplasty.

The purpose of our study was to assess and compare the psychometric properties of the Danish QuickDASH and PRWE in a group of patients with total wrist arthroplasty (TWA) with regard to construct validity, reproducibility, internal consistency, responsiveness and floor/ceiling effects.

MATERIAL AND METHODS

Study populations

In Group 1, we included consecutive patients operated with a third generation TWA at Gentofte Hospital or at Rigshospitalet, Denmark, during the 1999-2013 period. We evaluated the patients with the DASH or QuickDASH questionnaires: eight Universal (Integra Life Sciences Corp., Plainsboro, NJ, USA) and 96 Remotion (SBI Inc., Morrisville, PA, USA). Two patients were excluded because they did not attend the 12-month follow-up examination. This group (102 patients) was used for the assessment of the construct validity, internal consistency, floor/ceiling effects and responsiveness of the QuickDASH (**Figure 1**).

Group 2 consisted of a subset of Group 1: we included only the patients from Gentofte Hospital. This group was used for the general assessment of the PRWE and of the reproducibility of the QuickDASH (**Figure 1**).

There were 69 females and 33 males in Group 1, and 41 females and 22 males in Group 2. The mean age was 59.5 (29-83) years in Group 1 and 58.8 (31-83) in Group 2. There were 57 rheumatoid patients versus 45 non-rheumatoid patients in Group 1 and 29 rheumatoid versus 34 non-rheumatoid patients in Group 2.

Clinical design

In Group 1, we used data collected at the 12-month follow-up after TWA to evaluate construct validity, internal consistency, and floor/ceiling effects in a cross-sectional study. The responsiveness of the QuickDASH was calculated in a prospective cohort study using data collected preoperatively and at a 1-year follow-up.

Group 2 were entered into a cross-sectional study to evaluate the PRWE and into a test-retest trial to evaluate the reproducibility of both questionnaires. The

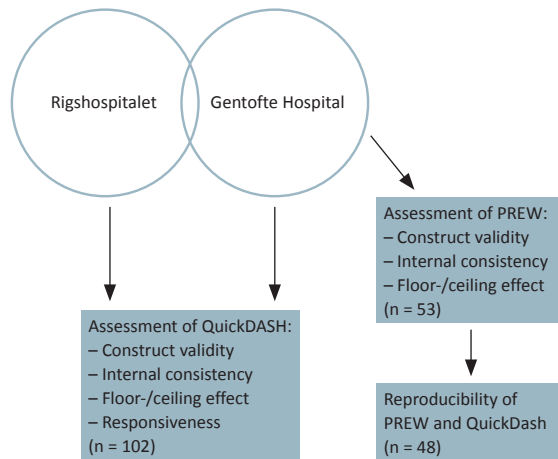
ORIGINAL ARTICLE

1) Clinic for Hand Surgery
2) Informatics Statistical Consulting

Dan Med J
2014;61(11):A4939


FIGURE 1

Flow chart illustrating how the two patient groups, treated at Rigshospitalet and Gentofte Hospital were selected. In Group 1, two patients had been excluded due to lack of response at the one-year follow-up. In Group 2, the number of patients was reduced from 53 to 48 in the test-retest trial due to lack of second response.



PRWE = Patient-rated Wrist Evaluation questionnaire.

QuickDASH = Quick Disabilities of Arm Shoulder and Hand questionnaire.

questionnaires were sent to the patients' private addresses by surface mail with a request to return it within a week and without informing them about the intention to retest. Six days after reception of the answer, a second questionnaire was sent in which we explained our wish to assess its reproducibility. A total of 53 returned the first questionnaire. One of these had insufficient answers to calculate a QuickDASH score, and we did not send her a second questionnaire. Four other patients did not return the second questionnaire. Thus, we had 48 sets of responses for the test-retest. The mean interval between the responses was 14.1 days (range: 6-29 days).

Instruments and measurements

The QuickDASH consists of six items concerning the ability to perform activities of daily living (ADL), two concerning social and work ability, one concerning pain and two concerning other symptoms. Each item has five response options (scored 1-5) which are used to create a summative score ranging from 0 (no disability or symptoms) to 100 (maximal disability or symptoms). If more than one item is missing in the QuickDASH, a score cannot be calculated.

The PRWE consists of eight items concerning ADL, two concerning social and work ability and five concerning pain. They are grouped into two sections: pain and function. Each item is rated on a Likert scale from 0 to 10 producing a summative score for each section. The total wrist score is calculated by adding the function

score divided by two plus the pain score, and ranges from zero (no disability or symptoms) to 100 (maximal disability or symptoms). If an item is missing, it is replaced with the mean score of the subscale. Both questionnaires were translated into Danish according to the Guillemin guidelines [5-7].

Reproducibility expresses to which extent scores can be reproduced in a test-retest.

Construct validity indicates the correlation between the measurements and theoretical considerations. To assess construct validity, we formulated three hypotheses a priori. Firstly, there should be a moderate, negative correlation between the scores and grip strength – the latter being a good indicator of hand function – but not a high correlation, considering the individual variance across patients related to their age, sex and body size. Secondly, we postulated a moderate, positive correlation with pain, because good function implies a low degree of pain. However, we did not expect a very high correlation since the questionnaire is not intended to simply measure pain. Thirdly, we postulated a weak or no correlation of the scales with wrist motion, knowing that even fused wrists are consistent with acceptable hand function.

For the testing of our hypotheses, we measured grip strength with the JAMAR (Sammon Preston Inc., Bolingbrook, IL, USA). We used a visual analogue scale (VAS) for evaluation of "general level of pain throughout the day". To express motion, we used the total dorsal/palmar wrist motion which was measured with a goniometer.

Floor and ceiling effects show the proportion of individuals who achieve the highest or lowest possible numeric value of a score and are considered present when more than 15% of the individuals achieve these values [8]. A ceiling or floor effect indicates that the measurement instrument cannot be used for the entire continuum of patients.

Cemented remotion total wrist arthroplasty.



Internal consistency measures to which extent the different items that propose to measure the same general construct tend to produce similar scores, i.e. whether there is general internal agreement between the items.

Responsiveness is the ability of a scale to measure a meaningful or important change in a clinical state, e.g. how it responds to treatment.

Statistical analysis

Correlations for construct validity were evaluated with Spearman's rho, values ± 0.8 to ± 1.0 indicating a very strong relationship, ± 0.6 to ± 0.8 a strong relationship, ± 0.4 to ± 0.6 a moderate relationship, ± 0.2 to ± 0.4 a weak relationship, and ± 0.0 to ± 0.2 a very weak or no relationship [9].

Internal consistency was assessed with Cronbach's alpha. Scales are considered to be internally consistent if Cronbach's alpha is between 0.7 and 0.9 [10]. Values higher than 0.9 might indicate item redundancy.

Responsiveness was expressed with the standardised response mean (SRM) and the effect size (ES) [11]. We considered values between 0 and 0.2 as "trivial", between 0.2 and 0.5 as "small", between 0.5 and 0.8 as "moderate" and higher than 0.8 as "large".

Reproducibility was expressed with Spearman's rho and the intraclass coefficient (ICC3) with the same interval definitions as mentioned.

Rheumatoid patients typically have multiple joint involvement to a higher extent than non-rheumatoid patients. We made separate analyses for these diagnostic subgroups in order to evaluate any bias. To compare the scores between rheumatoid and non-rheumatoid cases, we used the Wilcoxon rank-sum test.

The level of significance was set at $p < 0.05$.

Trial registration: not relevant.

RESULTS

For 17 of 308 testings (5.5%), the QuickDASH score could not be calculated, whereas all of the 100 PRWE scores could be calculated. Rheumatoid patients scored significantly higher according to the QuickDASH in Group 1, but not in Group 2. **Table 1** shows the result of the testing of the hypotheses for construct validity. Reproducibility, internal consistency and responsiveness are listed in **Table 2**. There were no statistically significant differences between the psychometric properties of the questionnaires in rheumatoid and non-rheumatoid cases. There was a very high correlation between the QDASH and the PRWE scores (Spearman's rho = 0.90). The scatter plot in **Figure 2** demonstrates that the scores are very similar, but not exactly numerically equivalent. The QuickDASH scores are approximately five points

TABLE 1

Construct validity of QuickDASH and PRWE according to three a priori formulated hypotheses concerning the correlation of the scores with grip strength, pain and wrist motion.

Spearman's rho	QuickDASH (p-value)	PRWE (p-value)
Versus grip strength	-0.56 (< 0.001)	-0.50 (< 0.001)
Versus pain	0.54 (< 0.001)	0.59 (< 0.001)
Versus motion	-0.10 (0.13)	0.22 (0.14)

PRWE = Patient-rated Wrist Evaluation questionnaire.

QuickDASH = Quick Disabilities of Arm Shoulder and Hand questionnaire.

TABLE 2

Reproducibility, responsiveness, internal consistency and floor-ceiling effect of the questionnaires.

	QDASH	PRWE
<i>Reproducibility, Group 2</i>		
Spearman's rho (p-value)	0.90 (< 0.001)	0.91 (< 0.001)
ICC, mean (range)	0.91 (0.85-0.95)	0.92 (0.87-0.96)
<i>Responsiveness, Group 1</i>		
SRM ^a	1.06	NA
ES	1.07	NA
<i>Internal consistency, Cronbach's alpha</i>		
Group 1	0.96	NA
Group 2	NA	0.97
<i>Floor-ceiling effect, % of total</i>		
Group 1:		
Floor	4	NA
Ceiling	0	NA
Group 2:		
Floor	NA	4
Ceiling	NA	0

ES = effect size; ICC = intraclass coefficient; NA = not applicable; PRWE = Patient-rated Wrist Evaluation questionnaire; QuickDASH = Quick Disabilities of Arm Shoulder and Hand questionnaire; SD = standard deviation; SRM = standardised response mean.

a) Mean score (\pm SD) was 55.2 (\pm 18.3) preoperatively and 35.4 (\pm 23.7) post-operatively. Difference in score was 19.6 (\pm 18.5).

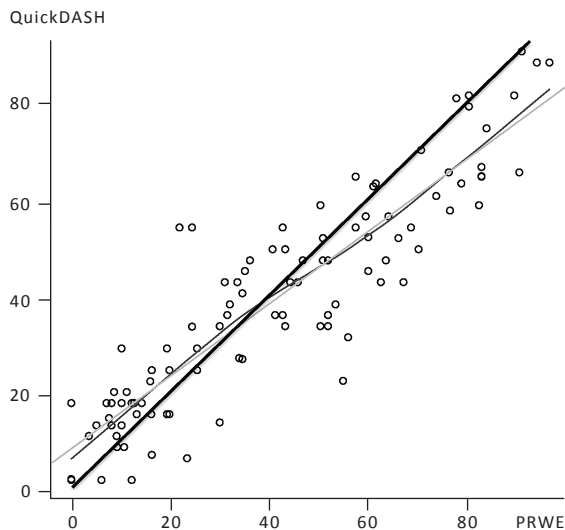
higher in the lower end of the scales (low disability), while they are approximately ten points lower in the higher end (high disability).

DISCUSSION

Earlier reports have demonstrated the validity of the QuickDASH for assessment of patients with shoulder, elbow and basal thumb joint arthroplasty [3] and the validity of the PRWE for assessment of basal thumb joint arthroplasty [12]. Our study indicates excellent psychometric properties of the questionnaires when applied to patients with TWA. The a priori formulated hypotheses concerning construct validity were confirmed. Reproducibility was very high. Cronbach's alpha indicated a strong internal consistency and possibly redundancy of items. There were no floor/ceiling effects. Responsive-


FIGURE 2

Scatter plot showing the QuickDASH- and the PRWE-scores in patients with total wrist arthro-plasty. The curved line is the LOESS (local regression) line, the thin solid line is the regression line and the thick solid line is the line of equivalency. Dots located on the line of equivalency represent patients whose QuickDASH and PRWE-scores are equal. Dots below this line indicate a QuickDASH-score that is lower than the PRWE-score for a given patient and vice-versa. The trend indicated by the LOESS-line is that the QuickDASH scores are approximately five points higher than the PRWE-scores in the lower end of the scales, whereas they are approximately ten points lower in the higher end.



PRWE = Patient-rated Wrist Evaluation questionnaire.
QuickDASH = Quick Disabilities of Arm Shoulder and Hand questionnaire.

ness to treatment was high according to the QuickDASH. It may be argued that we could have chosen to measure internal consistency, construct validity and floor/ceiling effects of the QuickDASH in Group 2, as we did for the PRWE, but we chose to take advantage of a larger available sample. In the systematic review of the measurement properties of the QuickDASH and its cross cultural adaptations performed by Kennedy et al [2], the studies were assessed with a recently described method: consensus-based standards for the selection of health measurement instruments (COSMIN). The studies with the best methodological quality showed high Cronbach's alpha values for internal consistency (0.92-0.94) and ICC values that ranged from 0.90 to 0.94. Hypothesis testing was evaluated in nine studies with a range of overall methodological quality: one excellent, six fair and two poor. Correlations were in the expected magnitude and direction: high for target construct (pain, function), moderate for work disability measures and low for mental health measures. Coefficients for the correlation with pain were 0.64 to 0.73. Several studies found acceptable SRM/ES after treatment of known efficacy (0.58-1.77), which indicates that the Quick DASH is sensitive to varying amounts of change. Thus, the reported psychometric

properties were generally in agreement with our findings, although they were not investigated in patients with TWA.

At present, no systematic review of the psychometric properties of the PRWE is available, and it is far beyond the scope of this study to do so with the COSMIN method, but we have identified a number of relevant papers. **Table 3** shows a summary of the internal consistency, reproducibility and responsiveness in these studies. A weak correlation with wrist motion and a moderate correlation with grip strength were found in patients with tendon interposition arthroplasty [12, 16, 18]. Correlations with VAS scores were moderate [14, 16, 17]. Fairplay et al found a high internal consistency (Cronbach's alpha = 0.96) and reproducibility (Spearman's rho = 0.93) in 63 patients with chronic wrist or hand pain [15]. These figures also are consistent with our findings in TWA patients.

One weakness of our study is that we were unable to assess the responsiveness of the PRWE since the Danish questionnaire was not available when we started sampling data in 2003. The surgical procedure itself is infrequent to an extent that it is necessary to collect data throughout several years in order to obtain sufficiently large numbers. Apart from this flaw, our analysis shows very similar psychometric properties for the two questionnaires. The fact that the scores produced by the scales were numerically very close is unexpected, because the QuickDASH is generally considered a generic upper limb instrument, whereas the PRWE is considered a specific wrist evaluation instrument. Notwithstanding the important fact that the psychometric properties of the two questionnaires did not differ between rheumatoid and non-rheumatoid patients, the scores of the rheumatoid patients in Group 1 were higher than the scores of the non-rheumatoid patients, which means that they had a higher grade of disability.

This must be interpreted in the light of the fact that the rheumatoid patients generally have multiple joint involvement. The difference was not significant in Group 2, which might be attributed to a smaller and hence an underpowered sample. It is also interesting and unexpected that the scores of both scales correlated equally with the general pain level, expressed on a visual analogue scale, because the PRWE contains a very detailed section with five specific pain questions, whereas the QDASH only has two general pain questions. This confirms that the level of pain is important, but not crucial for the construct measured and that multiple questions regarding pain may be superfluous and may contribute to item redundancy. The lack of correlation of the scores with mobility is in agreement with the study of Murphy et al 2003 [19] that failed to demonstrate any difference in DASH or PRWE scores between TWA and total wrist



TABLE 3

Psychometric properties of different versions of the PRWE in published studies.

Reference	Version	Diagnosis	Cases, n	Internal consistency, Cronbach's alpha	Reproducibility	Responsiveness
MacDermid et al, 1998 [13]	English	Wrist fractures	101	–	ICC > 0.90	–
MacDermid et al, 2000 ^b	English	Wrist fractures	59	–	–	SRM = 2.27
Kim & Kang, 2013 [14]	Korean	Wrist fractures	63	0.94	ICC = 0.96	ES = 0.94 SRM = 0.98
Wilcke et al, 2009 ^b	Swedish	Wrist fractures	99/49 ^a	0.94-0.97	Spearman's rho = 0.99	ES = 1.3 SRM = 1.4-1.7
Mellstrand et al, 2011 ^b	Swedish	Wrist injuries	124	0.97	ICC = 0.93	SRM = 1.29
Fairplay et al, 2012 [15]	Italian	Chronic wrist pain	63	0.96	Spearman's rho = 0.93	–
Imaeda et al, 2010 ^b	Japanese	Various wrist conditions	117/50 ^a	0.95	ICC = 0.92	ES = 1.9 SRM = 1.6
John et al, 2008 [12]	German	CMC1-arthritis after interposition	103/51 ^a	0.97	ICC = 0.86	–
Hemelaers et al, 2008 [16]	German	Wrist fractures	44	0.98	ICC = 0.94	–
Wong et al, 2007 ^b	Chinese (Hong Kong)	Wrist injuries	47	0.78-0.93	–	–
Schønnemann et al, 2013 [17]	Danish	Wrist fractures	60	0.94	ICC = 0.88	ES = 0.62
Present study	Danish	Total wrist arthroplasty	52/48 ^a	0.97	ICC = 0.92 Spearman's rho = 0.91	ES = 1.07 SRM = 1.06

ES = effect size; CMC = carpometacarpal; ICC = intraclass coefficient; PRWE = Patient-rated Wrist Evaluation questionnaire; SD = standard deviation; SRM = standardised response mean.

a) The second figure refers to a subset of cases used in a test-retest.

b) Reference available from corresponding author.

fusion (TWF) in a group of patients with generalised arthritis.

In our study, we were unable to assess the *criterion validity* of the scales. Criterion-related validity is based on evidence that shows the extent to which the scores of the instrument are related to a criterion measure or gold standard. The challenge is that there is no readily available gold standard: The authors and the institutions that developed the QuickDASH state that its criterion validity has not been tested because of the absence of a gold standard measure for the concept of upper-limb disability and hand function [20]. Nor could we make a direct assessment of the *respondent burden*, defined as the time, energy, and other demands placed on the patients to whom the instruments are administered because our patients answered the questionnaires unattended. However, the low number of missing items and the number of responses in the test-retest indicate an acceptable respondent burden and feasibility.

Future research must include other methods for evaluating scaling properties, like the Rasch analysis, not the least to assess unidimensionality and possible item redundancy suggested by the very high Cronbach's alpha.

CORRESPONDENCE: Michel E.H. Boeckstyns, Kløverbakken 11, 2830 Virum, Denmark. E-mail mibo@dadlnet.dk

ACCEPTED: 20 August 2014

CONFLICTS OF INTEREST: none. Disclosure forms provided by the authors are available with the full text of this article at www.danmedj.dk

LITERATURE

1. Beaton DE, Wright JG, Katz JN. Development of the QuickDASH: comparison of three item-reduction approaches. *J Bone Joint Surg* 2005;87:1038-46.
2. Kennedy CA, Beaton DE, Smith P et al. Measurement properties of the QuickDASH (disabilities of the arm, shoulder and hand) outcome measure and cross-cultural adaptations of the QuickDASH: a systematic review. *Qual Life Res* 2013;22:2509-47.
3. Angst F, Goldhahn J, Drerup S et al. How sharp is the short QuickDASH? A refined content and validity analysis of the short form of the disabilities of the shoulder, arm and hand questionnaire in the strata of symptoms and function and specific joint conditions. *Qual Life Res* 2009;18:1043-51.
4. MacDermid JC. Development of a scale for patient rating of wrist pain and disability. *J Hand Ther* 1996;9:178-83.
5. Guillemin F, Bombardier C, Beaton D. Cross-cultural adaptation of health-related quality of life measures: literature review and proposed guidelines. *J Clin Epidemiol* 1993;46:1417-32.
6. Herup A, Merser S, Boeckstyns M. Validation of questionnaire for conditions of the upper extremity. *Ugeskr Læger* 2010;172:3333-6.
7. Schønnemann JO. A comparison between two types of osteosynthesis for distal radius fractures using validated Danish outcome measures. Aarhus: Faculty of Health Sciences, University of Aarhus, 2010.
8. Terwee CB, Bot SD, de Boer MR et al. Quality criteria were proposed for measurement properties of health status questionnaires. *J Clin Epidemiol* 2007;60:34-42.
9. Chung MK. Correlation coefficient. In: Salkin NJ, ed. *Encyclopedia of measurement and statistics*. London: Sage Publications, 2007:189-201.
10. Tavakol M, Dennick R. Making sense of Cronbach's alpha. *Int J Med Edu* 2011;2:53-5.
11. Husted JA, Cook RJ, Farewell VT et al. Methods for assessing responsiveness: a critical review and recommendations. *J Clin Epidemiol* 2000; 53:459-68.
12. John M, Angst F, Awiszus F et al. The patient-rated wrist evaluation (PRWE): cross-cultural adaptation into German and evaluation of its psychometric properties. *Clin Exp Rheumatol* 2008;26:1047-58.
13. MacDermid JC, Turgeon T, Richards RS et al. Patient rating of wrist pain and disability: a reliable and valid measurement tool. *J Orthop Trauma* 1998;12:577-86.
14. Kim JK, Kang JS. Evaluation of the Korean version of the patient-rated wrist evaluation. *J Hand Ther* 2013;26:238-43.
15. Fairplay T, Atzei A, Corradi M et al. Cross-cultural adaptation and validation of the Italian version of the patient-rated wrist/hand evaluation questionnaire. *J Hand Surg Eur* 2012;37:863-70.

16. Hemelaers L, Angst F, Drerup S et al. Reliability and validity of the German version of "the Patient-rated Wrist Evaluation (PRWE)" as an outcome measure of wrist pain and disability in patients with acute distal radius fractures. *J Hand Ther* 2008;21:366-76.
17. Schonemann JO, Hansen TB, Soballe K. Translation and validation of the Danish version of the Patient Rated Wrist Evaluation questionnaire. *J Plast Surg Hand Surg* 2013;47:489-92.
18. MacDermid JC, Wessel J, Humphrey R et al. Validity of self-report measures of pain and disability for persons who have undergone arthroplasty for osteoarthritis of the carpometacarpal joint of the hand. *Osteoarthritis Cartil* 2007;15:524-30.
19. Murphy DM, Khoury JG, Imbriglia JE et al. Comparison of arthroplasty and arthrodesis for the rheumatoid wrist. *J Hand Surg Am* 2003;28:570-6.
20. Kennedy C, Beaton DE, Solway S et al, eds. DASH outcome measure user's manual, third edition. Toronto, Canada: Institute for Work & Health, 2011.