VALIDITY OF THE DISABILITY OF THE ARM, SHOULDER AND HAND PATIENT-REPORTED OUTCOME MEASURE (DASH) AND THE QUICKDASH WHEN USED IN DUPUYTREN'S DISEASE

Authors: JN Rodrigues[1], W Zhang[2], BE Scammell[2], PG Russell[3]*, I Chakrabarti[4]*, S Fullilove[5]*, D Davidson[6]*, TRC Davis[2]

Affiliations: [1]Nuffield Department of Orthopaedics, Rheumatology and Musculoskeletal Sciences, University of Oxford, Botnar Research Centre, Nuffield Orthopaedic Centre, Windmill Road, Oxford, OX3 7HE

[2]Division of Rheumatology, Orthopaedics and Dermatology, University of Nottingham & Nottingham University Hospitals NHS Trust, Queen's Medical Centre, Derby Road, Nottingham, UK

[3]Pulvertaft Hand Centre, Royal Derby Hospital, Uttoxeter New Road, Derby, UK

[4]Rotherham General Hospital, Moorgate Road, Rotherham, UK

[5]Derriford Hospital, Derriford Road, Plymouth, UK

[6]St John's Hospital at Howden, Howden Road South, Livingston, UK

*These authors contributed equally to this work

Corresponding author: JN Rodrigues, Nuffield Department of Orthopaedics, Rheumatology and Musculoskeletal Sciences, University of Oxford, Botnar Research Centre, Nuffield Orthopaedic Centre, Windmill Road, Oxford, OX3 7HE

07929296069, j.n.rodrigues@doctors.org.uk

Keywords: Dupuytren's contracture; Dupuytren's disease; Patient reported outcome measures; DASH; QuickDASH; hand function; correlation; agreement; validity

Acknowledgements: JNR received educational support from a Scholarship from the National Institute for Health and Care Excellence (NICE) during this project.

Mrs Audrey Parks, Research & Postgraduate PA at Pulvertaft Hand Centre, Derby, provided administrative support to this work. Mr Dariush Nikkhah, Specialty Registrar in Plastic Surgery, read the manuscript and provided invaluable feedback.

Conflicts of interests: none

Funding statement: This work was supported by a BSSH Research Fellowship, Nottingham Hospitals Charity and Nottingham Orthopaedic Walk.

Ethical approval: This study was a minor element of a larger service evaluation project studying treatment outcome in Dupuytren's disease. In keeping with UK National Research Ethics Service guidance, it is exempt from ethical approval. Approval as service evaluation was prospectively obtained from all centres involved. Local information governance was prospectively obtained from all centres involved, including Caldicott Guardian approval where required.

VALIDITY OF THE DISABILITY OF THE ARM, SHOULDER AND HAND PATIENT-REPORTED OUTCOME MEASURE (DASH) AND THE QUICKDASH WHEN USED IN DUPUYTREN'S DISEASE

SUMMARY

This study investigated aspects of the validity and reliability of the 30-item Disability of the Arm, Shoulder and Hand patient-reported outcome measure (DASH) and its relationship with the shorter 11-item QuickDASH in patients with Dupuytren's disease.

Seven hundred and fifty nine DASH questionnaires were studied, covering pre and postoperative patients undergoing different treatments for Dupuytren's disease. Items related to pain rose early after treatment before returning to baseline, suggesting that studying pain is relevant during postoperative recovery. Across all 759 sets of responses, the QuickDASH agreed closely with the DASH. In exploratory factor analysis, the DASH was not unidimensional, questioning the validity of the DASH summary score in Dupuytren's disease.

Further validation of existing PROMs for use in Dupuytren's disease is needed. These data suggest that pain is a relevant symptom to study during postoperative recovery following treatment for Dupuytren's disease.

Level of Evidence: III

INTRODUCTION

The use of patient-reported outcome measures (PROMs) in healthcare has been promoted in the UK (Darzi, 2008, Department of Health, 2010); international standards exist for their study (Mokkink et al., 2010). Several PROMs have been used to evaluate Dupuytren's disease; the 30-item Disability of the Arm, Shoulder and Hand tool (DASH) is the most popular (Ball et al., 2013). However it has been suggested that the DASH may not be valid for use in Dupuytren's disease (Beaudreuil et al., 2011, Packham, 2011), as it does not correlate closely with angular deformity (Engstrand et al., 2009, Jerosch-Herold et al., 2011, Zyluk and Jagielski, 2007); neither does the QuickDASH (Budd et al., 2011). Furthermore, both include items that assess pain whereas it is claimed that Dupuytren's disease may not be painful (Beaudreuil et al., 2011). Other groups suggest that pain may be present (Hueston, 1963, Rodrigues et al., 2014, von Campe et al., 2012) and treatment-related pain may affect postoperative recovery.

Much of the data describing the validity and reliability of the DASH was obtained from mixed cohorts involving upper limb conditions widely accepted as painful (Kennedy et al., 2011). Other than the recent publication of the secondary analysis of a randomised controlled trial (Forget et al., 2014), there is limited data describing the DASH's validity and reliability in Dupuytren's disease specifically.

Other PROMs that have been used to assess Dupuytren's disease (Ball et al., 2013), include the Michigan Hand Questionnaire (MHQ) (Chung et al., 1998), the Patient Evaluation Measure (PEM) (Macey et al., 1995), the Unité Rhumatologique des Affections de la Main scale (URAM) (Beaudreuil et al., 2011), and the QuickDASH (Beaton et al., 2005) . In a study of patients with a range of hand conditions, the DASH took longer to complete than the PEM, but was quicker than the MHQ (Dias et al., 2008). Patients contributing to research, service evaluation or audit might be asked to complete more than one outcome measure. For example, a specific PROM

and a generic measure to assess health-related quality of life, such as the EuroQol 5 D (EQ5D) (Herdman et al., 2011), may be required to facilitate cost effectiveness analysis (NICE, 2008). As a result, using PROMs that are quicker for the patient to complete may be more convenient and facilitate higher response rates.

The QuickDASH comprises 11 of the 30 items in the DASH, and should be quicker to complete. However, it has not been used extensively in Dupuytren's disease (Ball et al., 2013).

Consensus-based standards for the selection of health status measurement instruments (COSMIN) have been developed (Mokkink et al., 2010). These define different aspects of the validity of PROMs.

- Content validity assesses whether the items that comprise a PROM are an adequate reflection of what is trying to be measured. It involves assessing the relevance and comprehensiveness of the items in a PROM.
- Construct validity examines hypotheses about the PROM. Such hypotheses may relate to its structural validity (internal relationships between items), hypothesis testing (assessing its relationship with other PROMs) and differences between groups (cross-cultural validity).
- Internal consistency is the interrelatedness of the items within a PROM. This assumes that all of the items that contribute to a summary score actually assess the same underlying entity, or factor (e.g. impairment of structures in the hand versus restriction of function involving the shoulder), i.e. they are 'unidimensional'.
- Criterion validity tests a PROM against a 'gold standard'. The only accepted methodology for this is the comparison of a shortened PROM against the long version (e.g. the QuickDASH against the DASH).

- Responsiveness is the ability to detect change over time. This differs from 'validity', in that responsiveness assesses a change score, whereas validity assesses a single time point score.

This cross sectional study assessed aspects of content validity, construct validity and reliability of the DASH in Dupuytren's disease, and studied its relationship with the QuickDASH.

METHODS

Patient recruitment and data collection

The data presented in this study were gathered as part of larger service evaluation. Patient recruitment took place between September 2011 and April 2013. Exclusion criteria were cognitive impairment preventing informed consent and refusal of invitation to participate.

The inclusion criteria were primary or recurrent Dupuytren's disease and either patients awaiting fasciectomy or dermofasciectomy at one UK hand surgery centre or patients available for assessment at five UK hand surgery centres 1 year or 5 years (+/- 2 months) after their surgery when the primary author (JR) was available.

Patients in the first inclusion criterion group, who were participating in an observational cohort study, were recruited at a routine preadmission clinic visit prior to surgery. Those who were eligible and consented to participate completed the DASH in the clinic prior to surgery. These patients were also sent questionnaires for completion by post at 3 weeks, 6 weeks and 1 year after surgery. Patients who were scheduled for surgery to the left and right hand at different times during the study recruitment period were eligible for recruitment twice. This happened on four occasions.

Patients in the second inclusion criterion group, who were participating in a cross sectional study of postoperative outcome, were invited to participate with a letter explaining the project and inviting them to participate on a voluntary basis, with a fixed stipend offered to cover travel expenses. A single surgeon (JR) assessed those who consented to participate in a special clinic. The assessment included collection of demographic data and completion of the 30-item DASH questionnaire.

Angular measurement: total passive extension deficit (TPED)

Patients who completed PROMs in a clinic (as opposed to completion by post – which was the case for all 3 and 6 week postoperative measurements) had the

passive extension deficit of the treated digit assessed by a single examiner. Total passive extension deficit (TPED) was calculated by adding the measured passive extension deficits of the metacarpophalangeal joint and proximal interphalangeal joint while the other joints of the digit were passively flexed. The distal interphalangeal joint is rarely treated in our practice, and was not readily assessable with the model of goniometer used in the study. The measurement thus minimised the influence of dynamism, but is likely to have underestimated the total active extension loss (Rodrigues et al., 2014).

Content validity: relevance of pain questions

The relevance of items assessing pain was assessed by extracting and analysing responses to question 24 of the DASH (which assesses pain, and is question 9 of the QuickDASH) and question 25 of the DASH (which assesses pain during specific activity; it is not in the QuickDASH) at different time points. It was hypothesised that if pain items were relevant, they would change significantly through the recovery period.

Construct validity and reliability

The structural validity of the DASH was investigated by studying the internal relationships of the 30 different items in the DASH (how they related to each other) to assess whether the tool is “unidimensional”. When used as instructed by the developer, the DASH generates a single summary ‘DASH score’, using all of the 30 items. This is in contrast to other tools such as the Michigan Hand Questionnaire, which generates several summary scores for different areas. For the single DASH score to be valid, all items contributing to the score should measure, or ‘reflect’, the same underlying entity or ‘factor’, in this case upper limb function, i.e. the tool should be unidimensional (Mokkink et al., 2010). To evaluate whether the DASH is unidimensional, exploratory factor analysis (EFA) was employed (Mokkink et al., 2010). EFA analyses the relationship between items when completed by different people to identify underlying latent factors that explain the variance in scores; the

differences seen between individuals across a population. Some of the relevant concepts involved are defined in Table 1. It was expected that the responses obtained would have a tendency towards low scores and so not fit a normal distribution. This was examined by calculating the kurtosis and skewness for items. If the responses were not normally distributed, then logarithmic transformation of all items would be performed (Pallant, 2010) and their distributions then reassessed prior to factor analysis. EFA may be performed using different statistical methods. In this study, principal axis factoring was used to extract latent factors that were being reflected by the DASH's items, and the number of factors extracted was determined and confirmed by using two different accepted techniques (scree plots and parallel analysis, see Table 1) (Cattell, 1966, Horn, 1965, Patil et al., 2007). If the DASH is unidimensional, then there should only be one factor extracted. Cronbach's alpha was calculated to assess internal consistency. However, this must be interpreted with caution if unidimensionality has not been confirmed.

Relationship between the DASH and QuickDASH

The DASH summary score was calculated using the standard formula provided:

$$DASH=((a/b)-1) \times 25$$

Where "a" is the sum of the scores for the responses completed (each response could be scored between one and five), and "b" is the number of responses the patient completed.

The QuickDASH summary score was calculated by extracting the answers to the relevant 11 questions. The score was calculated using the same formula as for the DASH, only with these eleven items.

Parametric analyses were used to compare the summary scores of the DASH and the QuickDASH as both have virtually continuous scales (each scored 0-100), and the sample comprised a large number of independent observations. This is in keeping with the central limit theorem (Norman, 2010). Pearson's correlation coefficients were calculated between the total scores for the DASH and the

QuickDASH for a) the total sample and b) for different time point subgroups. If the relationship between the QuickDASH and the DASH was not absolute and did not lie on the line of equality (i.e. the correlation coefficient was less than 1, the maximum possible correlation coefficient), then agreement was also studied. Agreement was assessed by calculating 95% limits of agreement, using Bland-Altman analysis of the difference between the QuickDASH and the DASH (Bland and Altman, 1986). The responsiveness was studied by calculating the effect size (mean change in score / standard deviation of baseline score).

Handling of incomplete responses

If more than three of the 30 responses are missing (i.e. fewer than 27/30 responses provided), then the DASH cannot be calculated (Kennedy et al., 2011). If more than one response of the eleven is missing (i.e. fewer than 10/11 responses provided), then the QuickDASH cannot be calculated (Kennedy et al., 2011). If either occurred, then that questionnaire was excluded from the study. Consequently, some of the remaining questionnaires had some missing data (up to 3/30 responses missing for the DASH, or 1/11 missing for the QuickDASH). As the study of the relationship between the DASH and QuickDASH used summary scores, this was of no consequence for that analysis. In the EFA, all such questionnaires were included for analysis, but with missing responses excluded pairwise.

Approvals

This study was a minor element of a larger service evaluation project studying treatment outcome in Dupuytren's disease. In keeping with UK National Research Ethics Service guidance, it is exempt from ethical approval. Approval as service evaluation was prospectively obtained from all centres involved. Local information governance was prospectively obtained from all centres involved, including Caldicott Guardian approval where required.

RESULTS

Demographics

768 DASH questionnaires were received. These described the preoperative or postoperative assessment of 527 different procedures. Nine cohort study questionnaires were incomplete to the extent that calculation of a summary score was not possible based on the guidance issued with the DASH or the QuickDASH, and they were excluded from all analysis. Thus, 759 DASH questionnaires describing 527 procedures on 523 patients were analysed (this is represented graphically in the online appendix). The 527 procedures comprised 126 needle aponeurotomies (fasciotomies), 327 fasciectomies and 74 dermofasciectomies. The mean age at the time of assessment was 68 (range: 34 to 94) years; 403 of the 523 (77%) patients were men. The demographics of patients at time of completion of the 759 questionnaires are shown in Table 2. TPED measurements were made at the time of completion of the DASH scores in 522 of the 759 occasions (109 preoperative and 413 postoperative).

Content validity: relevance of pain questions

Question 24 of the DASH (which is question 9 of the QuickDASH) requires participants to rate pain experienced in the arm, shoulder or hand in the preceding week on a scale from 1 (no pain) to 5 (severe pain). This question was completed in 750 of the 759 questionnaires studied. The median score for question 24 was 2/5 (“mild” pain) for the total study. This was also the case when preoperative responses were studied alone. Sixty-eight patients provided answers to question 24 at each of the preoperative, 3 weeks and 6 weeks time points. When these responses were compared, there was a significant difference between them ($p=0.003$, repeated measures ANOVA test). Specifically, scores were lower at 6 weeks than at 3 weeks (Tukey’s multiple comparison test). Question 25 of the DASH (which is not part of the QuickDASH) rates pain when performing a specific activity. It was completed in 745 of 759 questionnaires. The median score overall was again 2/5, and this was

also the case for preoperative responses. The median score was 3/5 at 3-weeks postoperatively, falling back to 2/5 at 6 weeks postoperatively. Sixty-one patients provided answers to question 25 at each of the preoperative, 3 weeks and 6 week time points. Again, when these were compared, there was a significant difference between them (p=0.003, repeated measures ANOVA test). Scores were higher at 3 weeks than preoperatively and were lower at 6 weeks than at 3 weeks (Tukey's multiple comparison test).

Construct validity and reliability

Across the subgroup that also had angular deformity measured, the DASH showed weak correlation with TPED (Pearson's *r* = 0.30 (95% CIs: 0.22 to 0.38). The QuickDASH also correlated weakly with TPED 0.29 (Pearson's *r* = 95% CIs: 0.21 to 0.37).

After logarithmic transformation, 29 of the 30 DASH items had normal distributions. Most correlation coefficients between log transformed DASH items were over 0.3. Some correlation between the items is required for EFA, as it studies their inter-relationships. In the EFA, the presence of two major factors was suggested based on all common tests for determining factor numbers (see online appendix, confirmed by parallel analysis) (Cattell, 1966, Horn, 1965, Patil et al., 2007). Hence the DASH was not truly 'unidimensional': its 30 items are likely to reflect two factors rather than one, which is what might be expected given that all DASH items are combined into a single score that is supposed to reflect 'upper limb function'. Thus, combining all 30 to make a single DASH score may not be appropriate. These two underlying factors that were identified in the EFA explained 57.5% and 5.3% of variance respectively. The results of the EFA are shown in Table 3, with the prerequisite tests in the footnotes (Kaiser, 1974). The outputs from an EFA, called factor loadings, are the correlation coefficients between each item in the DASH questionnaire and the mathematical derived factor generated by the analysis. Specific function items correlated well with Factor 1, whereas patient perception of impairment and

participation items (including pain-related items) generally correlated with Factor 2. The EFA was also rerun using raw, untransformed data, and generated the same pattern of results.

Cronbach's alpha for the DASH was 0.975. However its interpretation was limited by finding that the DASH is potentially not unidimensional. Cronbach's alpha for the QuickDASH was 0.933. Both results were consistent with there being redundancy of items within the scales.

Relationship between the DASH and QuickDASH

Across the entire study, the QuickDASH was higher than the DASH indicating apparently worse upper limb function (mean difference 1.6 (95%CIs: 1.3 – 1.8), paired t test. However the QuickDASH correlated very well with the DASH, (Pearson's *r* was 0.98 (95% CIs: 0.98 – 0.99), as shown in Chart 1).

Linear regression analysis was performed with the Y-intercept constrained to y=0, as the QuickDASH must equal zero if the DASH equals zero. Runs test confirmed that there was no significant deviation of the residuals from the model in the linear regression analysis (p=0.228). The slope for the relationship between the two was: QuickDASH = 1.054 x DASH.

Similar correlations were seen in separate preoperative, 3 week, 6 week, 1 year and 5 year follow-up subgroup analyses (Table 4).

The 95% limits of agreement between the QuickDASH and the DASH were -5.8 to +8.9 (Chart 2). As relatively few differences were outside the 95% limits of agreement for mean scores under 30 (those with good upper limb function), further Bland-Altman analyses were performed for scores considered asymptomatic (<15/100), and scores considered symptomatic (>15/100) by the DASH's creators (Kennedy et al., 2011). When the mean of the DASH and QuickDASH for a patient was 15 or less (asymptomatic upper limb function), the 95% limits of agreement were -3.3 to +5.5, and when the mean was over 15 (symptomatic upper limb function), they were -7.8 to +12.3.

In terms of responsiveness from preoperative measurement to one year postoperative measurement, the effect size for the DASH was 0.58, and the effect size for the QuickDASH was 0.64.

DISCUSSION

Present study findings

This study assessed several aspects of the validity and reliability of the DASH in Dupuytren's disease, and its relationship with the QuickDASH.

PROM items covering pain have been previously criticised in Dupuytren's disease, but our patients did report preoperative upper limb pain. This may have been due to other comorbid upper limb conditions rather than Dupuytren's disease, nonetheless, this would affect the overall function of their upper limb. Furthermore, pain levels rose early after surgery, and then fell back. This is important to capture, as postoperative pain may differ between treatments, and affect early recovery. Our data support the relevance of assessing pain when treating Dupuytren's disease, such that scales that do not measure pain (such as the Unité Rhumatologique des Affections de la Main (URAM) (Beaudreuil et al., 2011)) may not provide a comprehensive assessment as a result.

The DASH is not unidimensional, and its task-based items loading better with one construct, and patient perception items load best with another. Some other PROMs are multidimensional and generate separate subscale scores. Examples include the Michigan Hand Questionnaire, which has distinct subscales for entities that might be expected to behave as different 'constructs', such as function, pain and work (Chung et al., 1998). However, the DASH is designed to generate a single summary score, and its items were not necessarily selected to measure specific distinct constructs. As a result, the different constructs identified here may not be easily interpreted. The items loading with Factor 1 might do so as they are all task-based items. The items that load with Factor 2 might do so as they are patient perception items. However, an alternative explanation might be that items that load well with Factor 1 do so as they reflect activities involving the entire upper limb activities, and the items that load well with Factor 2 specifically reflect the patient's experience of Dupuytren's disease.

Either way, these data suggest that the DASH's single summary score may not be appropriate in Dupuytren's disease. However, Factor 2 accounted for relatively little variance, such that its presence may not completely preclude the use of the DASH and its summary score in Dupuytren's disease. If used in future studies, its selection as an endpoint for future studies should be carefully considered.

Interpreting whether agreement is adequate or not is a clinical decision (Bland and Altman, 1986). Given that the minimum detectable change at the 95% confidence level (MDC_{95}) for the DASH is around 13, and that the MDC_{95} of the QuickDASH is around 16 (Kennedy et al., 2011), then the 95% limits of agreement seen here are of similar magnitude to both tools' abilities to detect true change, though these MDCs have not been specifically confirmed in Dupuytren's disease. Therefore, we believe that the level of agreement seen here would support the use of the QuickDASH as an alternative to the DASH if either were considered appropriate in Dupuytren's disease. Furthermore, their responsiveness was similar.

Our large sample size allowed meaningful subgroup analysis, which demonstrated that close correlation between the two tools was seen at preoperative, early postoperative, and late outcome time points.

Limitations

There are limitations to this study. Our sample included some patients who provided more than one measurement. However, multiple measurements over time (preoperative, 3 weeks postop, 6 weeks post-operative and 1 year) do not constitute replicate measurements. They are best considered as independent assessments, as the patient's functional status is expected to be different at each time point, due to treatment of disease and progressive recovery.

We used logarithmic transformation in an attempt to normalise the positive skew and kurtosis encountered. Suggested methods for handling skewed data vary (Ferguson and Cox, 1993, Floyd and Widaman, 1995). However, similar results were obtained when the EFAs were run using raw, untransformed data.

Here, as in previous studies (Gummesson et al., 2006, Niekel et al., 2009), the QuickDASH and DASH were calculated from a single set of responses, on the DASH questionnaire's proforma. Intra-observer reproducibility, or test-retest reliability, is not observed when using a single set of responses. However, we believe that our chosen methodology was the most appropriate to fulfil the specific objective of this study. This was to determine whether the QuickDASH formula demonstrates acceptable criterion validity with the longer DASH formula for a given set of responses, in keeping with the COSMIN checklist (Mokkink et al., 2010).

Test-retest reliability has been shown to be consistently high in previous studies, as summarised in the user manual for the DASH and the QuickDASH (Kennedy et al., 2011). The question of the present study is: "For a given set of responses, does the QuickDASH formula give you the same summary score as the DASH formula?" If patients had completed the DASH and then the QuickDASH, then there would have been additional error due to intra-observer reproducibility. This would have affected the ability to answer the study question.

Relationship to existing literature

Although the DASH is the most commonly used Patient Reported Outcome Measure (PROM) tool in Dupuytren's research (Akhavani et al., 2015, Ball et al., 2013), Dupuytren's-specific outcome measures are available, for example the URAM scale (Beaudreuil et al., 2011) and the Southampton Dupuytren's score (Mohan et al., 2014). Preoperative Dupuytren's disease is considered painless, and pain is not assessed in the URAM (Beaudreuil et al., 2011). However, pain has been described as a symptom, particularly related to Dupuytren's nodules (Hueston, 1963, Rodrigues et al., 2014, von Campe et al., 2012). Our data demonstrated that pain was present preoperatively, increased at 3 weeks post-operative compared to baseline, returned to the preoperative level by 6 weeks postop, and did not disappear completely.

The QuickDASH was produced from the DASH using item reduction methodology (Beaton et al., 2005). The two showed good correlation in mixed cohorts of hand

surgery patients (Gummesson et al., 2006, Niekel et al., 2009). However, correlation is not appropriate for studying agreement (Bland and Altman, 1986, Bland and Altman, 1990, Schuck, 2004). We are not aware of any studies assessing agreement between the DASH and the QuickDASH in Dupuytren's disease specifically, with the technique recommended by Bland and Altman (Bland and Altman, 1990). Studying the strength of relationship between the two correlated measures, as has been done elsewhere, may conceal absolute differences in values between them. Such differences are unmasked when agreement is studied using other techniques, as has been done here (Bland and Altman, 1986, Bland and Altman, 1990).

Poor correlation between angular deformity and the DASH has been previously reported in Dupuytren's disease (Degreef et al., 2009, Engstrand et al., 2009, Jerosch-Herold et al., 2011, Zyluk and Jagielski, 2007). This has led to the suggestion that the DASH may not be valid for use in Dupuytren's disease (Packham, 2011), although the basis for this claim was a series of only seven patient interviews (Pratt and Byrne, 2009). Angular correction remains a very popular measure of outcome in the treatment of Dupuytren's disease (Trickett et al., 2014, Verheyden, 2015). However, such conclusions are dependent on angular deformity being a 'gold standard' for assessing Dupuytren's disease, allowing the assessment of criterion validity of outcome measures including PROMs. This is not appropriate: hand function is a latent construct that cannot be measured directly, and so no true 'gold standard' for it can exist (Mokkink et al., 2010).

Concerns about the absence of the expected unidimensionality for the DASH have been previously reported. Forget and collagues investigated the structural validity of the DASH as the secondary analysis of a randomised controlled trial of splinting in Dupuytren's disease (Forget et al., 2014). EFA was performed on preoperative and early postoperative DASH scores from 153 fasciectomies and dermofasciectomies randomised to receiving postoperative splinting or no postoperative splinting. It also

demonstrated a lack of unidimensionality for the DASH. However the different groups of DASH items that loaded well on different factors could not be as easily explained as they can in the present study. The present study evaluated preoperative, early postoperative and late postoperative outcomes of over 500 procedures, including needle aponeurotomies, and in contrast, our EFA showed a clear division between task items and patient perception items. Another study that investigated a heterogeneous cohort comprising a range of upper limb conditions (mainly affecting the shoulder) and only a small minority with Dupuytren's disease found the DASH items loaded on three distinct factors in EFA, and this was confirmed using other techniques classified as confirmatory factor analysis (Franchignoni et al., 2010). As with the present study, patient perception items 22 to 30 loaded separately from the main factor, but they also found that of the other items, those relating to manual function loaded distinctly from those assessing shoulder functions. Given that the equivalent of the present study's Factor 1 was two separate factors in a mixed cohort including shoulder patients suggests that Factor 1 is not a clean and reliable indicator of hand function in Dupuytren's disease. Our data further question the structural validity of the DASH in Dupuytren's disease, and by association, the QuickDASH. Due to this, we believe that it is neither necessary nor appropriate to subject the DASH to further analyses such as Rasch analysis (Tesio, 2003). Instead, other PROMs that also assess pain, such as the MHQ or the PEM, may be more appropriate for the study of Dupuytren's disease.

Summary

This study supports the assessment of pain when studying recovery from Dupuytren's surgery. The QuickDASH show acceptable agreement with the full DASH. However, the DASH may not be structurally valid for use in Dupuytren's disease, and further study of existing PROMs for use in Dupuytren's disease is needed.

REFERENCES

ir L. A review of the classification of Dupuytren's disease. J Hand Surg Brit Eur. 2015, 40: 155-65.

Ball C, Pratt AL, Nanchahal J. Optimal functional outcome measures for assessing treatment for Dupuytren's disease: a systematic review and recommendations for future practice. BMC Musculoskelet Disord. 2013, 14: 131.

Beaton DE, Wright JG, Katz JN. Development of the QuickDASH: comparison of three item-reduction approaches. J Bone Joint Surg Am. 2005, 87: 1038-46.

Beaudreuil J, Allard A, Zerkak D et al. Unite Rhumatologique des Affections de la Main (URAM) scale: development and validation of a tool to assess Dupuytren's disease-specific disability. Arthrit Care Res. 2011, 63: 1448-55.

Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. Lancet. 1986, 1: 307-10.

Bland JM, Altman DG. A note on the use of the intraclass correlation coefficient in the evaluation of agreement between two methods of measurement. Comput Biol Med. 1990, 20: 337-40.

Budd HR, Larson D, Chojnowski A, Shepstone L. The QuickDASH score: a patient-reported outcome measure for Dupuytren's surgery. J Hand Ther. 2011, 24: 15-20.

Cattell RB. The Scree Test for the Number of Factors. Mutlivar Behav Res. 1966, 1: 245-76.

Chung KC, Pillsbury MS, Walters MR, Hayward RA. Reliability and validity testing of the Michigan Hand Outcomes Questionnaire. J Hand Surg Am. 1998, 23: 575-87.

Darzi A. High quality care for all: NHS Next Stage Review final report. London: TSO, 2008.

http://www.dh.gov.uk/en/Publicationsandstatistics/Publications/PublicationsPolicyAndGuidance/DH_085825 (27th June 2012).

448 Degreef I, Vererfve PB, De Smet L. Effect of severity of Dupuytren contracture on
449 disability. Scand J Plast Recons. 2009, 43: 41-2.
450 Department of Health. Equity and Excellence: Liberating the NHS. London 2010.
451 http://www.dh.gov.uk/en/Publicationsandstatistics/Publications/PublicationsPolicyAnd
452 Guidance/DH_117353 (26th June 2012).
453 Dias JJ, Rajan RA, Thompson JR. Which questionnaire is best? The reliability,
454 validity and ease of use of the Patient Evaluation Measure, the Disabilities of the
455 Arm, Shoulder and Hand and the Michigan Hand Outcome Measure. J Hand Surg
456 Brit Eur. 2008, 33: 9-17.
457 Engstrand C, Boren L, Liedberg GM. Evaluation of activity limitation and digital
458 extension in Dupuytren's contracture three months after fasciectomy and hand
459 therapy interventions. J Hand Ther. 2009, 22: 21-6.
460 Ferguson E, Cox T. Exploratory Factor Analysis: A Users' Guide. Int J Select Assess.
461 1993, 1: 84-94.
462 Floyd FJ, Widaman KF. Factor Analysis in the Development and Refinement of
463 Clinical Assessment Instruments. Psychol Assessment. 1995, 7: 286-99.
464 Forget NJ, Jerosch-Herold C, Shepstone L, Higgins J. Psychometric evaluation of
465 the Disabilities of the Arm, Shoulder and Hand (DASH) with Dupuytren's contracture:
466 validity evidence using Rasch modeling. BMC Musculoskelet Disord. 2014, 15: 361.
467 Franchignoni F, Giordano A, Sartorio F, Vercelli S, Pascariello B, Ferriero G.
468 Suggestions for refinement of the Disabilities of the Arm, Shoulder and Hand
469 Outcome Measure (DASH): a factor analysis and Rasch validation study. Arch Phys
470 Med Rehabil. 2010, 91: 1370-7.
471 Gummesson C, Ward MM, Atroshi I. The shortened disabilities of the arm, shoulder
472 and hand questionnaire (QuickDASH): validity and reliability based on responses
473 within the full-length DASH. BMC Musculoskelet Disord. 2006, 7: 44.
474 Herdman M, C. G, Lloyd A et al. Development and preliminary testing of the new
475 five-level version of EQ-5D (EQ-5D-5L). Qual Life Res. 2011, 20: 1727-36.

Horn JL. A Rationale and Test for the Number of Factors in Factor Analysis. Psychometrika. 1965, 30: 179-85.

Hueston JT. Principles of management. In: Hueston JT (Ed.) *Dupuytren's contracture*. London, E &S Livingstone, 1963: 76-94.

Jerosch-Herold C, Shepstone L, Chojnowski A, Larson D. Severity of contracture and self-reported disability in patients with Dupuytren's contracture referred for surgery. J Hand Ther. 2011, 24: 6-10.

Kaiser HF. An index of factorial simplicity. Psychometrika. 1974, 39: 31-6.

Kennedy C, Beaton D, S S, S M, Bombardier C. *The DASH and QuickDASH Outcome Measure User's Manual*, Third ed. Toronto, Institute for Work and Health, 2011.

Macey AC, Burke FD, Abbott K et al. Outcomes of hand surgery. British Society for Surgery of the Hand. J Hand Surg Brit Eur. 1995, 20: 841-55.

Mohan A, Vadher J, Ismail H, Warwick D. The Southampton Dupuytren's Scoring Scheme. J Plast Surg Hand Surg. 2014, 48: 28-33.

Mokkink LB, Terwee CB, Knol DL et al. The COSMIN checklist for evaluating the methodological quality of studies on measurement properties: a clarification of its content. BMC Med Res Methodol. 2010, 10: 22.

NICE. Guide to the methods of technology appraisal. London: NICE, 2008. http://www.nice.org.uk/aboutnice/howwework/devnicetech/technologyappraisalprocessguides/?domedia=1&mid=B52851A3-19B9-E0B5-D48284D172BD8459 (26th June 2012).

Niekel MC, Lindenhovius AL, Watson JB, Vranceanu AM, Ring D. Correlation of DASH and QuickDASH with measures of psychological distress. J Hand Surg Am. 2009, 34: 1499-505.

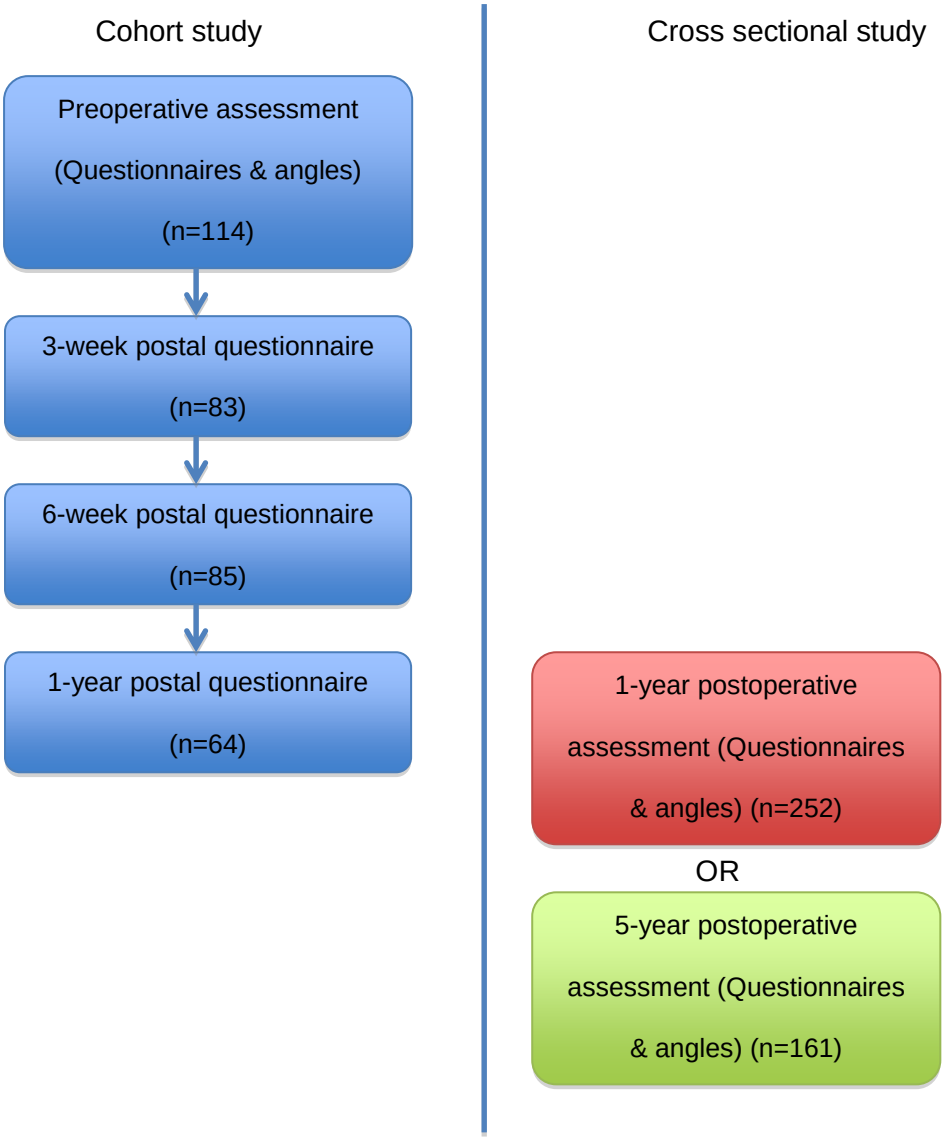
Norman G. Likert scales, levels of measurement and the "laws" of statistics. Adv Health Sci Educ Theory Pract. 2010, 15: 625-32.

503 Packham T. Clinical commentary in response to: Severity of contracture and self-
504 reported disability in patients with Dupuytren's contracture referred for surgery. J
505 Hand Ther. 2011, 24: 12-4.
506 Pallant J. Part 3: Preliminary analyses: Manipulating the data. *SPSS Survival*
507 *Manual*, 4th ed. Maidenhead, Open University Press, 2010.
508 Patil VH, Singh SN, Mishra S, Donovan TD. Parallel Analysis Engine to Aid
509 Determining Number of Factors to Retain [Computer Software]. 2007.
510 http://smishra.faculty.ku.edu/parallelengine.htm (2nd August 2014).
511 Pratt AL, Byrne G. The lived experience of Dupuytren's disease of the hand. J Clin
512 Nurs. 2009, 18: 1793-802.
513 Rodrigues JN, Zhang W, Scammell BE, Davis TR. Dynamism in Dupuytren's
514 contractures. J Hand Surg Brit Eur. 2014, 40: 166-70.
515 Rodrigues JN, Zhang W, Scammell BE, Davis TR. What patients want from the
516 treatment of Dupuytren's disease -- is the Unité Rhumatologique des Affections de la
517 Main (URAM) scale relevant? J Hand Surg Brit Eur. 2014, 40: 150-4.
518 Schuck P. Assessing reproducibility for interval data in health-related quality of life
519 questionnaires: which coefficient should be used? Qual Life Res. 2004, 13: 571-86.
520 Tesio L. Measuring behaviours and perceptions: Rasch analysis as a tool for
521 rehabilitation research. J Rehabil Med. 2003, 35: 105-15.
522 Trickett RW, Savage R, Logan AJ. Angular correction related to excision of specific
523 cords in fasciectomy for Dupuytren's disease. J Hand Surg Brit Eur. 2014, 39: 472-6.
524 Verheyden JR. Early outcomes of a sequential series of 144 patients with
525 Dupuytren's contracture treated by collagenase injection using an increased dose,
526 multi-cord technique. J Hand Surg Brit Eur. 2015, 40: 133-40.
527 von Campe A, Mende K, Omaren H, Meuli-Simmen C. Painful nodules and cords in
528 Dupuytren disease. J Hand Surg Am. 2012, 37: 1313-8.
529 Zyluk A, Jagielski W. The effect of the severity of the Dupuytren's contracture on the
530 function of the hand before and after surgery. J Hand Surg Brit Eur. 2007, 32: 326-9.

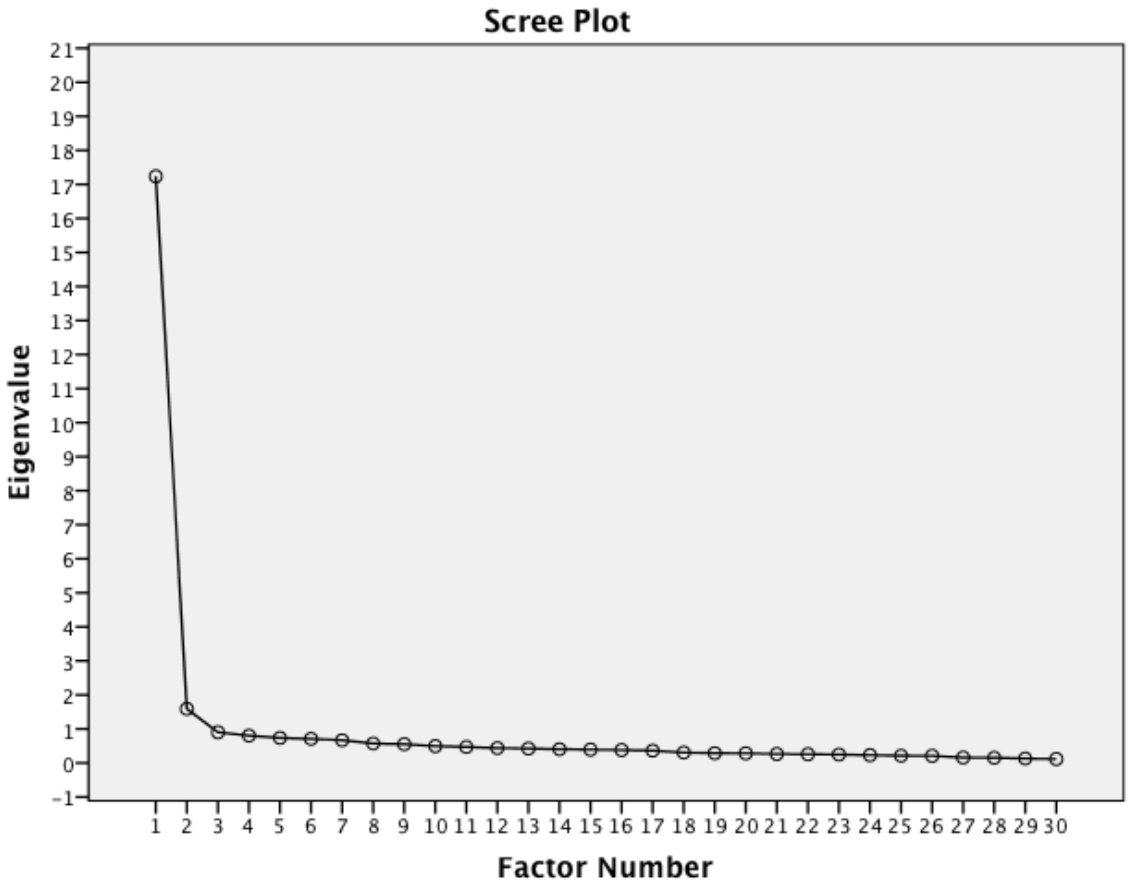
531
532

533

Appendix 1: Flow diagram demonstrating time points from which complete questionnaires were received



Appendix 2: Scree plot of exploratory factor analysis for log transformation of DASH items



The curve flattens from factor 3 onwards. Based on Cattell's test, this supports the extraction of factors 1 and 2. This was confirmed using parallel analysis, and only these two factors had Eigenvalues over one.

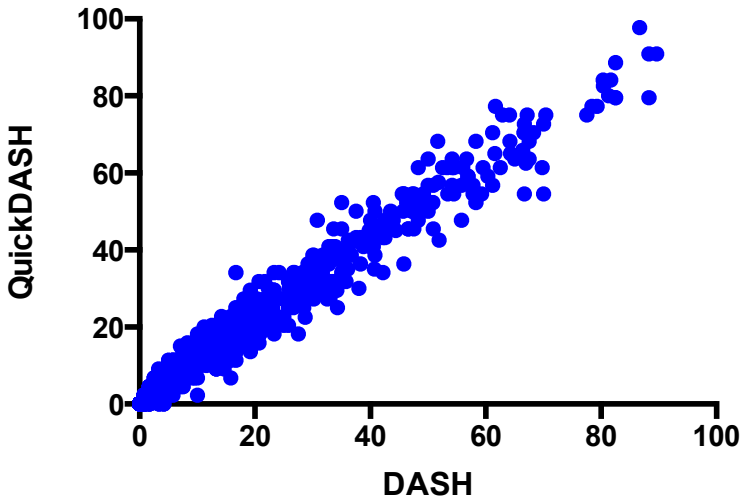


Chart 1: Scatterplot of QuickDASH versus DASH (n=759).

Chart 2: Bland-Altman plot of total sample, with 95% limits of agreement shown as dotted lines

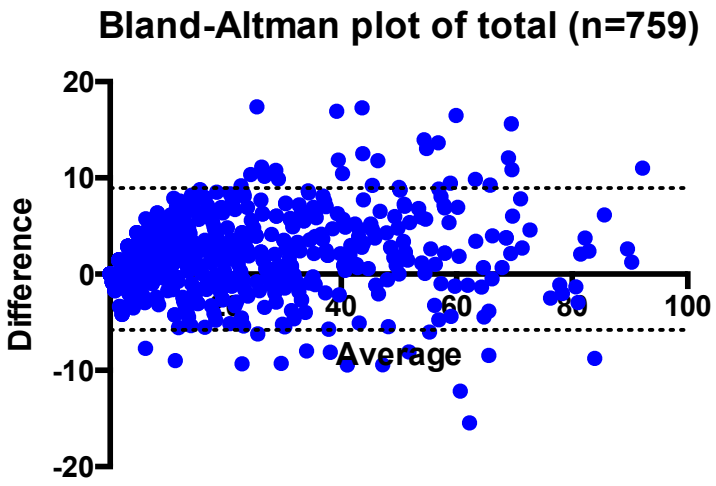


Table 1: Data handling and relevant statistical concepts

Aspect Studied	Method(s) Used	Concept(s)	Description
Content validity	Cohort study of pain items	Comprehensiveness & Relevance	Content validity considers whether PROM items are relevant to what is being measured, and whether the scale overall is comprehensive.
Distribution of item responses	Kurtosis & skewness	Kurtosis	Defines the sharpness of the peak of a distribution of data
		Skewness	Defines the amount of asymmetry of a distribution of data
Construct validity (structural validity)	Exploratory Factor Analysis	Unidimensionality	All items contributing to a summary score reflect the same underlying factor (some PROMs generate more than one summary score, each describing a different subscale of the PROM)
		Factor	Factor analysis aims to describe the variation in measured items that correlate with each other in terms of fewer unobserved 'factors'
		Principal axis factoring	A form of factor analysis in which factors are extracted based on common variance, rather than total variance.
		Eigenvalue	Describes the amount of the total variance that is explained by a particular factor
		Catell's scree plot	Determines the number of factors to extract. All potential factors are plotted in order of Eigenvalues (see Chart 1). The turning point where the connecting line flattens sharply is the point at which the last significant factor has been passed, as further factors represent a flat level of background noise (Catell, 1966).

		Parallel analysis	Determines the number of factors to extract. A second set of Eigenvalues is generated, but which are based entirely on chance. All factors in the real model with Eigenvalues greater than their counterparts in the chance model are significant and are then extracted. Avoids the risk of bias that exists with scree plot interpretation (Horn, 1965).
		Kaiser-Meyer-Olkin statistic	Assesses sampling adequacy. It lies between 0-1, describing the proportion of variance that is common variance. The minimum acceptable level for analysis being 0.6, and >0.9 being described as 'marvellous' (Kaiser, 1974).
		Bartlett's test of sphericity	Assesses whether an identity matrix would result if correlations between the items included were studied, i.e. whether the correlation between all of the DASH items is zero. Some correlation between items is needed for EFA. A significant result is achieved if the data is suitable for EFA.
		Factor loading	The output of EFA. Described how closely an item correlates with a factor.
		Rotation	Presents the factor loadings in a manner that makes interpretation more straightforward, resulting in a pattern matrix. Only possible when more than one factor is extracted. When done, the majority of the items may each show strong loading with one of the factors, a situation described as simple structure.
Reliability	Cronbach's alpha	Internal consistency	Studies the inter-relatedness between items within a scale. However, it is dependent on the scale being unidimensional and reflective
Relationship between DASH and QuickDASH	Bland-Altman plots	Agreement	Studies the relationship between two variables that are expected to correlate. In such circumstances, reporting correlation is common, but may not be appropriate.

1 Table 2: Subgroups and demographics of completed questionnaires

Subgroup	Number of complete questionnaires (incompletes)	Procedure type	Mean age at assessment (range)	Proportion male
Total	759 (9)	130 aponeurotomies 494 fasciectomies 135 dermofasciectomies	68 (34-94)	592/759 (78%)
Preoperative	114 (3)	N/A	67 (34-90)	88/114 (77%)
3 weeks postop	83 (0)	2 aponeurotomies 57 fasciectomies 24 dermofasciectomies	67 (34-90)	66/83 (80%)
6 weeks postop	85 (3)	2 aponeurotomies 62 fasciectomies 21 dermofasciectomies	68 (34-90)	68/85 (80%)

1 year postop	316 (3)	105 aponeurotomies 168 fasciectomies 43 dermofasciectomies	68 (35-91)	249/316 (79%)
5 years postop	161 (0)	19 aponeurotomies 126 fasciectomies 16 dermofasciectomies	71 (39-89)	121/161 (75%)

2

3

Table 3: Exploratory factor analysis output of log-transformed DASH items (Pattern matrix with two-factor solution and oblimin rotation)

[The pattern coefficients may be considered similar to correlation coefficients between the item and the factors was derived in EFA. They may range from -1 (perfect inverse correlation), through 0 (no correlation), to +1 (perfect correlation)]

Item number	Item question	Pattern coefficient with factor 1	Pattern coefficient with factor 2
12	Difficulty changing a light bulb overhead	**0.86**	-0.05
14	Difficulty washing your back	**0.84**	-0.09
7	Difficulty doing heavy household chores	**0.84**	0.05
5	Difficulty pushing open a heavy door	**0.82**	-0.06
13	Difficulty washing or blow drying hair	**0.82**	-0.03
9	Difficulty making a bed	**0.80**	0.03
16	Difficulty using a knife to cut food	**0.80**	0.01
11	Difficulty carrying a heavy object (>10 lbs)	**0.80**	0.04
8	Difficulty gardening	**0.80**	0.08
6	Difficulty placing an object on a shelf above head	**0.78**	-0.04
17	Difficulty with recreational activities that require little effort	**0.78**	-0.05
4	Difficulty preparing a meal	**0.77**	0.07
18	Difficulty with recreational activities in which force is taken through the limb	**0.73**	0.13
21	Difficulty with sexual activities	**0.72**	-0.11
19	Difficulty with recreational activities in which the arm moves freely	**0.71**	0.12
3	Difficulty turning a key	**0.69**	0.03
15	Difficulty putting on a pullover sweater	**0.68**	0.09
20	Difficulty managing transportation needs	**0.68**	0.02
10	Difficulty carrying a shopping bag/briefcase	**0.67**	0.15
2	Difficulty writing	**0.63**	0.02
1	Difficulty opening a tight or new jar	**0.61**	0.15

22	To what extent has your problem interfered with normal social activities?	**0.45**	**0.37**
23	To what extent has your problem interfered with work or other daily activities?	**0.44**	**0.42**
24	Rate your arm, shoulder or hand pain	-0.06	**0.90**
25	Rate your arm, shoulder or hand pain when performing a specific activity	-0.02	**0.87**
28	Rate the stiffness in your arm, shoulder or hand	0.08	**0.72**
27	Rate the weakness in your arm, shoulder or hand	0.17	**0.69**
29	How much difficulty have you had sleeping because of pain in the limb?	0.10	**0.62**
26	Rate the tingling in the arm, shoulder or hand	-0.04	**0.59**
30	Is this true: I feel less capable, confident or useful because of the limb problem?	0.29	**0.51**

N.B. Items have been sorted by size of factor loading. Large factor loadings (>0.3) are shown in bold.

The Kaiser-Meyer-Olkin statistic for the analysis was 0.97 and Bartlett's test of sphericity was highly statistically significant (Kaiser, 1974).

Table 4: Mean DASH and QuickDASH scores and their correlations, by time point

TIME POINT	MEAN DASH (95% CIs)	MEAN QuickDASH (95% CIs)	PEARSON's *r* (95% CIs)
Preoperative	27 (23-31)	28 (24-32)	0.98 (0.97-0.99)
3 week postoperative	37 (33-42)	41 (36-46)	0.97 (0.95-0.98)
6 week postoperative	22 (18-26)	24 (20-28)	0.98 (0.97-0.99)
1 year postoperative	11 (9-13)	12 (10-14)	0.99 (0.98-0.99)
5 years postoperative	11 (9-13)	12 (10-15)	0.97 (0.96-0.98)