Development and Validation of a Disease-Specific Questionnaire for Basal Joint Arthritis

Peter C. Noback, BA¹ Joseph M. Lombardi, MD¹ Mani Seetharaman, MD, MS¹ Donald H. Lee, MD²
Robert J. Strauch, MD¹ Melvin P. Rosenwasser, MD¹

¹Department of Orthopaedic Surgery, Trauma Training Center, Columbia University Medical Center, New York, New York

²Vanderbilt Orthopedic Institute, Hand and Upper-Extremity Center, Medical Center East, Nashville, Tennessee

Address for correspondence Melvin P. Rosenwasser, MD, Department of Orthopaedic Surgery, Trauma Training Center, PH-1164, 622 West 168th Street, New York, NY 10032 (e-mail: ttc@columbia.edu).

J Wrist Surg 2017;6:126–133.

Abstract

Keywords

- thumb
carpometacarpal
arthritis
- basal joint
- outcome instrument
- thumb disability
examination
- disease specific

Background The Disabilities of the Arm, Shoulder, and Hand (DASH) questionnaire is the most commonly used instrument to assess outcomes of basal joint arthritis (BJA). However, the DASH is subject to influence by the entire upper extremity.

Purpose This study aims to develop and validate a disease-specific questionnaire for BJA that would be more sensitive to changes in thumb function and pain, and correlate better with objective and subjective metrics.

Patients and Methods The thumb disability examination (TDX) was developed and 80 patients presenting with BJA at one of the two hospital-based hand clinics were enrolled in the validation study. At enrollment, subjects were given the TDX, DASH, and visual analog pain scale with activity (A-VAS) surveys. The strength was assessed. Patients receiving corticosteroid injection were seen for follow-up at 6 weeks and those who underwent surgery were seen between 3 and 6 months postoperatively. Both the groups were given the TDX, DASH, and A-VAS scales at follow-up.

Results In total, 65 subjects were included in the analysis. Average TDX completion time was 134.3 seconds. The TDX correlated more strongly with A-VAS scores at baseline than the DASH, but less strongly with tip-pinch measures. The TDX was more responsive to injection and surgical treatments for BJA than the DASH, yielding a larger effect size and standardized response mean, and was the only instrument to significantly correlate with changes in A-VAS.

Conclusion The TDX is a reliable instrument for assessing BJA treatment outcomes. It bears less of a burden on patients, is more responsive to symptomatic changes, and correlates better with most objective and subjective measures than the DASH.

Level of Evidence II, diagnostic.

Patient self-reported measures are commonly used to compare outcomes and quality of medical care.^{1,2} These instruments should ideally be practical for patients to complete, correlate predictably with functional outcomes, and be responsive to clinically significant changes in patients' symptoms.³ In studies examining the upper extremity, the most commonly used

outcome score is that of the Disabilities of the Arm, Shoulder, and Hand (DASH) questionnaire. The DASH is a global outcome instrument pertaining to all portions of the upper extremity that has been validated for use in multiple languages.^{1,4–10}

The DASH is composed of 30 items, 21 of which are related to function, 6 to the severity of symptoms, and 3 to

received

August 8, 2016

accepted

September 5, 2016

published online

October 14, 2016

Copyright © 2017 by Thieme Medical Publishers, Inc., 333 Seventh Avenue, New York, NY 10001, USA.
Tel: +1(212) 584-4662.

DOI <http://dx.doi.org/10.1055/s-0036-1593612>.
ISSN 2163-3916.

psychological factors.¹¹ The DASH has been compared with and tested against other instruments specific to diseases and joints of the upper extremity and has been found to be at least as responsive (i.e., able to accurately discern different magnitudes of symptomatic/functional change) as these questionnaires.^{3,7} However, disease-specific questionnaires for conditions of the hand, such as carpal tunnel syndrome and distal radius fractures, provide improved responsiveness and correlation with symptoms when compared with the DASH.³ Therapeutic or interventional research for specific conditions such as these may, therefore, benefit from the improved responsiveness of more specific surveys.

While basal joint osteoarthritis (BJA) of the thumb is the most frequently reported form of clinically impactful arthritis in the hand, and the most frequently surgically treated arthritic condition of the upper extremity, there are no commonly used and validated disease-specific questionnaires that target BJA.^{12–14} Currently, DASH scores are often used to make comparisons of outcomes between surgical and nonsurgical treatments for BJA.¹⁴ However, such scores are influenced by pathophysiological conditions or injuries relating to both upper extremities, calling into question the accuracy of study results that utilize DASH scores as endpoints.

Certain hand- and wrist-specific questionnaires, such as the Michigan Hand Questionnaire (MHQ), Australian/Canadian Hand Osteoarthritis Index (AUSCAN), and Patient Rated Wrist Evaluation (PRWE), have been compared with the DASH in terms of their ability to discriminate between severities of BJA.^{9,15} In these often cross-sectional comparisons, it has been found that hand-specific questionnaires have similar discriminatory abilities to the DASH with respect to BJA. Given the common usage of the DASH in research outside of BJA, it is not surprising that it thus remains the instrument of choice in BJA literature over these also valid yet less frequently utilized questionnaires.

The purpose of this study was to develop and validate a disease-specific questionnaire for BJA to improve accuracy when comparing treatment outcomes. We compared this questionnaire to the DASH and hypothesized that it would be more practical for patients, more sensitive to changes in thumb function and pain, and better correlate with objective and subjective metrics.

Materials and Methods

Instrument Development

The questionnaire, titled the thumb disability examination (TDX), was developed by physicians and researchers at a single institution. The development process consisted of item generation, item reduction, and pilot testing, followed by final item reduction.

Stage I: Item Generation

Disease-specific and global orthopedic outcome scales related to the upper extremity were reviewed and items from these scales related to the thumb were collected.^{11,16–21} The MHQ, DASH, and AUSCAN were primarily used to gather and conceive items related to tasks commonly hindered by BJA.

The PRWE and McGill Pain questionnaire were used as a reference to generate items that focused on thumb pain. Furthermore, 30 BJA patients presenting with stages I to IV of the Eaton–Little classification for BJA were interviewed in the clinic and asked to provide a list of activities and movements that were most hindered by their disease, as well as to discuss specific aspects of their thumb with which they were least satisfied.²² The senior author (M.P.R.) was subsequently consulted and potential items not found in reviewed scales were added based on this input. A total of 125 unique items were generated as a result of this process.

Stage II: Item Reduction

The initial list of items was reviewed by a panel of hand surgeons, hand therapists, and 20 BJA patients not included in the item generation. Surgeons and therapists were asked to remove items they believed to be easily influenced by other disabilities of the hand or wrist, items irrelevant to BJA, items that did not require force applied by the thumb, and items that were too specific to dominant hands. Two items that were initially deemed too specific to the dominant hand were retained to the next stage of development due to panelist's agreement that they were frequently hindered by BJA. BJA patient panelists were then asked to remove items that did not relate to their thumb, items that related to their thumb but were unaffected by their arthritis, and items that were unclear in their description. Members of the panel were then asked to group the remaining items into one out of six domains: function, pain, aesthetics, work performance, social influences, and satisfaction with the affected thumb. All the remaining 45 items fell into the domains of function, pain, and satisfaction. A preliminary questionnaire was then developed by the researchers (P. N., J. L., and M. R.) and reviewed by an epidemiologist experienced in survey creation.

Stage III: Pilot Testing, Final Instrument Production

A pilot testing program distributed the questionnaire to 50 BJA patients presenting with all stages of BJA. When items were left blank, patients were asked to provide a brief explanation of why. From items left blank, those that elicited reasoning of being inapplicable to patients, not performed by the affected hand, or too difficult to understand, were removed. In total, 35 items remained after this removal. Factor analysis was then initiated for a final round of item removal, which resulted in the removal of 15 items that were not isolated to one domain (function, pain, or satisfaction). For instance, analysis of a question assessing difficulty typing on a phone found that the response to this question had feedback that was not consistent with that of other functionality questions. Using this analytical technique ensured that each question was only associated with one domain.

The final form of the TDX was composed of 20 questions divided into three sections based on the described domains. The TDX is scored on a scale of 0 to 100, with a higher score indicating a greater degree of disability in the thumb. The scoring rubric for the TDX and the finalized questionnaire can be found in the **Supplementary Material** (available in the online version), respectively.

Instrument Reliability and Validation

Study Design

The validation process and sample sizes were based on the definitions and recommendations of the COSMIN group.²³ Over a 2-year period (November 2013–November 2015) the TDX and DASH were administered to 120 patients presenting at one of the two separate hospital-based hand clinics by one of the three research assistants/fellows. The DASH was chosen as a comparator instead of a hand-specific questionnaire because of its higher frequency of use in BJA literature as compared with scales such as the MHQ or AUSCAN.¹⁴ Furthermore, visual analog scale pain with activity (A-VAS) scores, range of motion, and strength measures were collected by a trained research assistant or fellow at the first visit. Inclusion criteria included a presentation with a chief complaint relating to BJA in either hand. The diagnosis was based on clinical information, plain radiographs, and physical tests including: the grind test, the lever test, and the metacarpophalangeal extension test.²⁴ Exclusion criteria included previous enrollment in a BJA registry coordinated by the two affiliated hospitals, age under 18 years, lack of English fluency, current or history of infection of the basal joint, or current active treatment for BJA with no management change. The study received approval from our institutional review board and was funded in part by the Orthopedic Science and Research Foundation.

Patients were divided into three study groups based upon the treatment. Those patients who were treated conservatively with either a thumb splint, nonsteroidal gels, or any combination of the two were placed into one study group, termed SNSG (splints/nonsteroidal gels). Those given a corticosteroid injection were placed in a second study group, termed IJ. Patients who were scheduled for surgery to treat the symptomatic joint were placed in a third group, termed SG. Treatments were determined by clinical need and not by the study protocol.

Patients in study group IJ were followed up with 6 weeks after injection and asked to fill out the TDX, DASH, and A-VAS scales. Patients in study group SG were followed up between 4 and 6 months after surgery and asked to fill out the TDX, DASH, and A-VAS scales. Any patients who were not seen for follow-up in groups IJ or SG were excluded from longitudinal analysis.

Subjective and Objective Measures

The DASH questionnaire and A-VAS are previously well described and were unmodified for our protocol.^{10,25} Grip strength was assessed by instructing patients to fully adduct their elbows, maintain a 90-degree angle between their forearm and upper arm, and actuate a Jamar dynamometer (TEC, Clifton, NJ) on setting II. Tip-pinch strength was assessed with identical arm position and actuation of a Jamar hydraulic pinch gauge with their affected thumb on top and index finger underneath. Each strength measure was conducted three times, and an average strength measure was then calculated.

Demographics

Baseline demographic and clinical characteristics of patients in all groups were analyzed and compared by one-way analysis of variance (ANOVA) testing.

Practicality

To assess the patient burden, the time to complete the TDX at the initial visit was measured in all patients. Completion time of the TDX at subsequent visits and the completion times of other instruments were not measured due to staff shortage.

Floor and Ceiling Effects

TDX floor and ceiling effects were assessed for the total score and for the three sections by calculating the fraction of patients who scored a minimal or maximum score, relative to the total number of patients. A maximal score was considered a score of 90 to 100, while a minimal score was considered a score of 0 to 10. If more than 10% of the participants achieved a maximal or minimal score, then this would be considered a ceiling or floor effect, respectively.

Internal Consistency

Cronbach α was calculated using TDX questionnaires completed by patients at their initial visit. This coefficient is a representation of the extent to which all of the items of a test measure the same latent variable. The value of the coefficient ranges from 0.0 to 1.0, with 1.0 indicating the highest correlation possible. A high reliability ($\alpha > 0.95$) may indicate the presence of redundant items.

Test-Retest Reliability

The TDX was administered to a group of 20 patients who elected to receive no treatment for their chief complaint of BJA. After 2 weeks of initial administration of the TDX, these patients were then asked to complete the questionnaire a second time. This period of time was considered long enough to avoid potential recall bias, but short enough to avoid progression of arthritis. From these scores, intraclass correlation coefficients (ICC) were calculated for the total score as well as the three sections. ICCs can range from 0.0 to 1.0, with 1.0 indicating the strongest reliability.²⁶

Construct Validity

Construct validity assesses if an instrument behaves expectantly (i.e., greater thumb disability results in a higher TDX score) and was assessed by calculating Pearson correlation coefficients between the TDX total and sectional scores, DASH, A-VAS, grip strength, and tip-pinch strength at baseline. Correlations were calculated collectively and between treatment groups. In the individual analysis of correlation, groups IJ and SG were treated as one to avoid diluting the sample size and limiting statistical analysis. This combined group is labeled as IJ + SG. Furthermore, correlations between the changes in TDX, DASH, and A-VAS scores from baseline to follow-up for the combined study group IJ + SG were also calculated. A Pearson coefficient with an r value from 0.0 to 0.25 was defined as poor, 0.26 to 0.50 as moderate,

0.51 to 0.75 as good, and 0.76 to 1.0 as excellent.²⁶ A negative correlation was defined in an inverse manner.

Responsiveness

TDX and DASH responsiveness were measured using IJ + SG scores at baseline and at follow-up. These were used to calculate effect size (ES) and standardized response mean (SRM) of the instruments as well as the sectional scores in the TDX.²⁷ ES is defined as the difference in means between baseline and follow-up divided by the standard deviation (SD) at baseline. SRM is defined as the difference in means between baseline and follow-up divided by the SD of the difference between the scores at these time points. ES and SRM values categorize the magnitude of the difference between instruments' scoring mechanisms and ranges of variation. These are defined as follows: < 0.20 is trivial, 0.21 to 0.50 is small, 0.51 to 0.70 is moderate, and > 0.71 is large.

A *p* value < 0.05 was considered statistically significant.

Results

Demographics

Over 2 years, 80 patients were enrolled in this study. Seven patients were excluded due to a partially completed DASH questionnaire, three due to a partially completed TDX questionnaire, and five due to incomplete follow-up. Thus, 65 patients were included in the analyses, for a total of 68 hands, with 3 patients being actively treated for bilateral BJA. There were a total of 23 males and 42 females, with an average age of 67.3 years. The study group SNSG was composed of 31 patients, IJ was composed of 33 patients, and SG was composed of

4 patients. SG patients were seen for follow-up at an average of 4.9 (SD: 0.9) months after surgery. Total TDX and sectional scores were normally distributed across and within all groups. Descriptive statistics and results of the ANOVA analysis of the differences across groups can be found in ►Table 1.

Practicality

Average time to complete the TDX at initial administration was 134.3 (SD: 26.4) seconds (range: 75.0–200.0 seconds).

Floor and Ceiling Effects

For the baseline total TDX scores, as well as the baseline scores of the function and pain sections, less than 10% of patients achieved a maximal or minimal score, meaning that any floor or ceiling effects were negligible. Nine (13.2%) hands (seven patients) achieved a score greater than 90 on the satisfaction section, signifying a ceiling effect in this domain.

Internal Consistency

The total TDX score, as well as its three sections, showed high internal consistency as assessed by Cronbach α , with values ranging from 0.79 for the section related to pain, to 0.93 for both the total TDX score and the section related to function. Cronbach α values can be found in ►Table 2 for all sections.

Test-Retest Reliability

Test-retest administrations to 20 patients found no statistically significant differences between initial and follow-up total TDX, function, pain, or satisfaction score (*p* > 0.05). The test-retest correlation coefficients for the total TDX, function, pain, and satisfaction sections were 0.97, 0.92, 0.88, and 0.98, respectively.

Table 1 Demographic and baseline descriptive statistics of 65 patients (68 hands) who underwent treatment for BJA, stratified by treatment group

	SG	IJ	SNSG	Differences between study groups	All subjects
	<i>n</i> = 4	<i>n</i> = 33	<i>n</i> = 31	<i>p</i> Value	<i>n</i> = 65
Age (y)	67.3 ± 7.8	68.2 ± 9.8	66.3 ± 9.8	0.90	67.3 ± 9.7
Gender, <i>n</i> (%)					
Male	2 (50)	19 (57.6)	21 (67.7)	0.82	23 (35.4)
Female	2 (50)	14 (42.4)	10 (32.3)	0.76	42 (64.6)
Dominant side affected, <i>n</i> (%)	2 (50)	18 (54.5)	15 (48.4)	0.90	35 (51.4)
Nondominant side affected, <i>n</i> (%)	2 (50)	15 (45.5)	16 (51.6)	0.90	33 (48.6)
DASH	51.3 ± 13.5	36.4 ± 16.0	30.5 ± 16.1	0.04 ^{ab}	34.7 ± 16.7
TDX	60.0 ± 20.0	58.6 ± 17.8	45.3 ± 17.4	0.04 ^{ab}	47.6 ± 18.1
A-VAS	7.6 ± 0.9	7.3 ± 1.9	6.38 ± 2.6	0.18	7.0 ± 2.3
R-VAS	2.3 ± 2.1	3.2 ± 2.4	2.9 ± 2.8	0.78	3.0 ± 2.6

Abbreviations: A-VAS/R-VAS, visual analog scale for pain with activity/at rest; BJA, basal joint arthritis; DASH, Disabilities of the Arm, Shoulder, and Hand score; IJ, study group of patients receiving injections; SG, surgical study group; SNSG, study group of patients receiving nonsteroidal gel and/or splint; TDX, thumb disability examination score.

Note: Values are expressed as mean ± standard deviation unless otherwise indicated. All counts and percentages are related to total number of subjects (*n* = 65), except for metrics related to the side affected, which are calculated with respect to total number of hands (*n* = 68).

^aStatistically significant difference between groups SG and SNSG.

^bStatistically significant difference between groups IJ and SNSG.

Table 2 Internal validity of the TDX and correlations of the TDX and DASH with subjective patient measures at baseline ($n = 65$)

	Cronbach α	Correlations with DASH and A-VAS					
		DASH			A-VAS		
		IJ + SG	SNSG	All subjects	IJ + SG	SNSG	All subjects
Total TDX ($n = 68$)	0.93	0.63 ^a	0.74 ^a	0.63 ^a	0.34 ^b	0.45 ^b	0.40 ^a
Function	0.93	0.60 ^a	0.73 ^a	0.66 ^a	0.25	0.33 ^b	0.31 ^a
Pain	0.79	0.46 ^a	0.47 ^a	0.44 ^a	0.35	0.47 ^a	0.40 ^a
Satisfaction	0.84	0.38 ^b	0.59 ^a	0.46 ^a	0.03	0.42 ^b	0.35 ^a
DASH ($n = 68$)		N/A	N/A	N/A	0.21	0.28	0.28 ^b

Abbreviations: DASH, Disabilities of the Arm, Shoulder, and Hand score; IJ + SG, study group of patients receiving either surgery or corticosteroid injection ($n = 37$ hands); n , the total number of hands; N/A, not applicable; SNSG, study group of patients receiving nonsteroidal gel and/or splint ($n = 31$ hands); TDX, thumb disability examination score.

Note: Correlations are expressed as Pearson correlation coefficients. Cronbach α is a measure of internal validity. Function, pain, and satisfaction: the three sections of the TDX.

^a p value < 0.01.

^b p value < 0.05.

Construct Validity

Pearson correlation coefficients between baseline TDX and sectional scores, DASH, and A-VAS can be found in ►Table 2. Pearson correlation coefficients between baseline TDX and sectional scores, DASH, grip strength, and tip-pinch strength can be found in ►Table 3. Of note, TDX scores were significantly correlated with DASH and A-VAS scores in all groups. The section that correlated best with the DASH was the pain section ($p = 0.66$, < 0.01). Total TDX score and pain section scores both correlated the best with A-VAS ($p = 0.40$, < 0.01). Total TDX score, the functional section score, and DASH significantly correlated with tip-pinch strength overall ($p = -0.29$, -0.33 , and -0.38 , respectively, $p < 0.05$).

Responsiveness

As seen in ►Table 4, total TDX, TDX sectional, A-VAS, and DASH scores all significantly changed from baseline to follow-up visit for group IJ + SG. The total TDX score had a larger ES and SRM compared with the DASH, indicating greater

responsiveness to clinical changes in BJA. Both the ES and SRM for the total TDX were rated as large, while those of the DASH as moderate. As seen in ►Table 5, the TDX score was the only instrument that significantly correlated with change in A-VAS ($p = 0.63$, < 0.01).

Discussion

Development and validation of new quantitative survey instruments have recently accelerated as patients, insurers, and hospitals have broadened their focus on outcome factors to include safety, patient satisfaction, and costs.^{1,28,29} The health care industry is transitioning to payment for quality, which is measured in part by these instruments. In hand surgery, outcomes traditionally were quantified by metrics such as strength and range of motion.³⁰ In BJA, outcomes related to function, pain, and overall satisfaction are equally as important.^{12,13} Currently, the DASH is the tool most often used to assess these domains in this patient subset.¹⁰ However, the

Table 3 Correlations of the TDX and DASH with objective patient measures

	Correlations with grip and tip-pinch strength					
	Grip strength			Tip-pinch strength		
	IJ + SG	SNSG	All subjects	IJ + SG	SNSG	All subjects
Total TDX ($n = 68$)	-0.13	0.16	-0.01	-0.24 ^a	-0.35	-0.29 ^a
Function	-0.15	0.14	-0.10	-0.25	-0.43 ^a	-0.33 ^a
Pain	-0.10	0.10	0.01	-0.18	-0.13	-0.15
Satisfaction	0.01	0.18	0.10	-0.08	-0.14	-0.12
DASH ($n = 68$)	-0.20	-0.26	-0.26	-0.27	-0.56 ^b	-0.38 ^a

Abbreviations: DASH, Disabilities of the Arm, Shoulder, and Hand score; IJ + SG, study group of patients receiving either surgery or corticosteroid injection ($n = 37$ hands); n , the total number of hands; N/A, not applicable; SNSG, study group of patients receiving nonsteroidal gel and/or splint ($n = 31$ hands); TDX, thumb disability examination score.

Note: Correlations are expressed as Pearson correlation coefficients. Function, pain, and satisfaction: the three sections of the TDX.

^a p value < 0.01.

^b p value < 0.05.

Table 4 Baseline and follow-up data of subjects treated with corticosteroid injection or surgery

	Baseline	Follow-up	Δ	p Value	ES	SRM
A-VAS ($n = 37$)	7.4 ± 1.9	4.8 ± 1.9	-2.41 ± 1.5	< 0.001	-1.3	-1.6
DASH ($n = 37$)	37.8 ± 16.4	29.3 ± 15.1	-8.9 ± 13.1	< 0.001	-0.54	-0.68
Total TDX ($n = 37$)	48.7 ± 17.3	30.3 ± 20.9	-18.0 ± 18.3	< 0.001	-1.04	-0.98
Function	42.3 ± 22.6	37.8 ± 23.8	-9.6 ± 18.0	< 0.001	-0.42	-0.53
Pain	47.1 ± 18.6	24.7 ± 18.8	-20.9 ± 18.6	< 0.001	-1.1	-1.1
Satisfaction	68.3 ± 18.1	42.9 ± 24.8	-26.1 ± 28.7	< 0.001	-1.4	-0.91

Abbreviations: A-VAS/R-VAS, visual analog scale for pain with activity/at rest; DASH, Disabilities of the Arm, Shoulder, and Hand score; ES, effect size; n , the total number of hands; IJ, study group of patients receiving injections; SG, surgical study group; SRM, standardized response mean; TDX, thumb disability examination score.

Note: Values are expressed as mean $\pm$ standard deviation unless otherwise indicated. For ES and SRM, < 0.20 is trivial, 0.21 to 0.50 is small, 0.51 to 0.70 is moderate, and > 0.71 is large. Negative values were defined similarly. Subjects assessed include those in study groups IJ ($n = 33$ hands) and SG ($n = 4$ hands). Follow-up data collection occurred at 6 weeks after injection for patients in group IJ, and between 3 to 6 months postoperatively for patients in group SG. Function, pain, satisfaction: the three sections of the TDX.

DASH cannot discriminate between changes caused by BJA and other upper-extremity conditions, such as carpal tunnel syndrome. In addition, patients suffering from contralateral hand conditions necessarily confound outcome scores when using the DASH to assess an intervention. Due to frequently occurring comorbid conditions, prior BJA studies have sought to use custom questionnaires to assess outcomes.³¹ However, these instruments are almost never tested or reproduced, making comparisons between studies impossible.^{32–35} Though the MHQ is a validated hand-specific questionnaire that was shown to be slightly more responsive than the DASH with respect to BJA, it is composed of 37 items that take an average of 8 to 20 minutes to complete, and thus imposes an even greater burden on patients than the DASH in terms of both item quantity and completion time.^{15,36,37}

A previous attempt at developing a BJA-specific questionnaire included a 10-item instrument called the Nelson Score, which was previously evaluated in a small subset ($n = 36$) of patients undergoing surgery for BJA and demonstrated moderate correlation with pain in the arthritic thumb.³⁸ Validation studies of

the DASH, Boston Carpal Tunnel Questionnaire, and MHQ were comparatively better powered. In addition, no correlation of subjective and objective measures was performed between the DASH and the Nelson Score. Properties of the DASH that are considered limiting, such as administration time, were also not formally assessed in the Nelson Score. Consequently, we sought to develop and validate a new disease-specific outcome questionnaire for BJA through a larger sample study to improve physician's ability to compare treatment outcomes for this frequently treated disease.

There are several limitations to this study. First, only the completion time for the TDX was measured in this population. While measuring completion time of the DASH would have allowed for a better comparison of each questionnaire's burden on patients, we believe our comparisons are still valid. Second, when assessing the responsiveness of the TDX and DASH, patients receiving corticosteroid injections for BJA, a treatment not always shown to be effective, were used. Because such injections are not definitively effective, responsiveness in the TDX and DASH may be underestimated in this

Table 5 Longitudinal responsiveness

	Correlation between changes from baseline to follow-up			
	A-VAS		DASH	
	Pearson	p Value	Pearson	p Value
Total TDX ($n = 37$)	0.63	< 0.01	0.60	< 0.001
Function	0.44	< 0.05	0.44	< 0.05
Pain	0.22	0.29	0.64	< 0.001
Satisfaction	0.37	0.07	0.48	< 0.001
DASH	0.21	0.33		

Abbreviations: A-VAS, visual analog scale for pain with activity; DASH, Disabilities of the Arm, Shoulder, and Hand score; ES, effect size; n , the total number of hands; IJ, study group of patients receiving injections; SG, surgical study group; SRM, standardized response mean; TDX, thumb disability examination score.

Note: Correlation of changes in total TDX Score and each of its three sections with changes in A-VAS and DASH from baseline to follow-up visit for patients in groups IJ + SG. IJ + SG, study group of patients receiving either surgery or corticosteroid injection ($n = 37$ hands). Correlations were assessed by Pearson correlation coefficients. Follow-up visits occurred at 6 weeks after injection for patients in group IJ and 3 to 6 months after surgery for patients in group SG. Function, pain, satisfaction: the three sections of the TDX.

subgroup. However, we believe both questionnaires' responsiveness would be affected in similar manners, and thus a comparison of ES and SRM between the questionnaires likely remains valid analysis. Finally, no comparisons were made to other questionnaires specific to the hand and wrist. This was not done due to anticipated questionnaire burden on patients, and due to the DASH being used much more frequently in BJA literature than questionnaires such as the MHQ, PRWE, or AUSCAN, despite some evidence showing that in certain settings these questionnaires may be more responsive than the DASH.³

Logistically, the TDX required a mean completion time of 134 seconds at the initial clinic visit. The DASH was previously found to have an average completion time of 7 minutes, a much longer time in comparison.³⁹ The reduced completion time of the TDX directly translates to a reduced questionnaire burden on patients, and in future research, could lead to lower patient exclusions due to incomplete questionnaires.

Our results also show that the TDX is a highly reliable instrument, with an overall test-retest correlation coefficient of 0.97, and each of the sections' correlation coefficients ranging from 0.88 to 0.98. Internal consistency of the entire instrument measured high with a Cronbach α value of 0.93, while the sectional scores ranged from 0.79 to 0.93. These α values indicate excellent internal consistency for items in each section and the TDX overall, and likely a lack of redundancy as well.

At baseline, the overall TDX and DASH demonstrated "good" correlation; a desirable result since higher correlation may indicate little added value to metrics assessed by the DASH. Though correlation with A-VAS was significant with both instruments, the overall TDX, and its pain section demonstrated stronger correlation with A-VAS. This correlative strength can be of great utility given that pain is an important impetus for treatment of BJA and can singularly define therapeutic success for many patients. The results regarding strength measures are interesting yet have been seen elsewhere: in prior BJA treatment studies, it is often the case that those studies that combine pain and strength measures often find conflicting results.^{40,41} Though exceptions exist, patients generally seek relief of pain over restoration of strength when undergoing treatment for BJA, and thus pain relief is cited as proof of a favorable outcome.

In patients receiving injections or surgical management, the TDX had a greater ES and SRM compared with the DASH, indicating a higher level of responsiveness to treatment. The overall TDX score change after treatment demonstrated statistically significant correlation with A-VAS score changes, unlike the DASH. Given that pain is an outcome of primary concern to patients, these results are encouraging with respect to the value of the TDX.

This study demonstrates the validation of the TDX and that it imposes less burden on patients, is more responsive to changes in BJA-related symptoms, and correlates better with most objective and subjective metrics relevant to BJA when compared with the DASH. In research seeking to compare treatment outcomes for this common condition, use of the TDX should be considered to provide more detailed and comparable results. Future research will compare the use of

the TDX to that of widely accepted and studied general health surveys such as the 36-Item Short Form Health Survey, as well hand-specific questionnaires, to further illustrate its psychometric properties and advantages over potentially burdensome or nondisease-specific instruments. Given the increasing importance of patient-reported outcome measures, instruments such as this are vital to ensure surgeons maintain an integral role in defining clinical excellence.^{42,43}

Note

This work was approved by the ethical review committee of the associated hospital. Data were collected at the Columbia University Medical Center and the Vanderbilt Orthopedic Institute.

Funding

This work was funded in part by a grant from the Orthopedic Science and Research Foundation.

Conflict of Interest

None.

Acknowledgments

The authors would like to thank Julie Daniels and Michael Silverman for their assistance in enrolling subjects in this study. They would also like to thank Lee Ann Finno for her administrative support on this study.

References

- Kamal RN, Kakar S, Ruch D, et al. Quality Measurement: A Primer for Hand Surgeons. *J Hand Surg Am* 2016;41(5):645–651
- Beaton DE, Bombardier C, Guillemin F, Ferraz MB. Guidelines for the process of cross-cultural adaptation of self-report measures. *Spine* 2000;25(24):3186–3191
- Beaton DE, Katz JN, Fossel AH, Wright JG, Tarasuk V, Bombardier C. Measuring the whole or the parts? Validity, reliability, and responsiveness of the Disabilities of the Arm, Shoulder and Hand outcome measure in different regions of the upper extremity. *J Hand Ther* 2001;14(2):128–146
- Veehof MM, Slegers 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). *J Hand Ther* 2002;15(4):347–354
- Gummesson C, Atroshi I, Ekdahl C. The disabilities of the arm, shoulder and hand (DASH) outcome questionnaire: longitudinal construct validity and measuring self-rated health change after surgery. *BMC Musculoskelet Disord* 2003;4:11. Doi: 10.1186/1471-2474-4-11
- Padua R, Padua L, Ceccarelli E, et al. Italian version of the Disability of the Arm, Shoulder and Hand (DASH) questionnaire. Cross-cultural adaptation and validation. *J Hand Surg [Br]* 2003;28(2):179–186
- Gay RE, Amadio PC, Johnson JC. Comparative responsiveness of the disabilities of the arm, shoulder, and hand, the carpal tunnel questionnaire, and the SF-36 to clinical change after carpal tunnel release. *J Hand Surg Am* 2003;28(2):250–254
- Jester A, Harth A, Wind G, Germann G, Sauerbier M. Disabilities of the arm, shoulder and hand (DASH) questionnaire: Determining functional activity profiles in patients with upper extremity disorders. *J Hand Surg [Br]* 2005;30(1):23–28
- MacDermid JC, Richards RS, Donner A, Bellamy N, Roth JH. Responsiveness of the short form-36, disability of the arm,

- shoulder, and hand questionnaire, patient-rated wrist evaluation, and physical impairment measurements in evaluating recovery after a distal radius fracture. *J Hand Surg Am* 2000;25(2):330–340
- 10 Beaton DE, Davis AM, Hudak P, McConnell S. The DASH (Disabilities of the Arm, Shoulder and Hand) outcome measure: what do we know about it now? *J Hand Ther Br*. 2001;6:109–118
 - 11 Hudak PL, Amadio PC, Bombardier C; The Upper Extremity Collaborative Group (UECG). Development of an upper extremity outcome measure: the DASH (disabilities of the arm, shoulder and hand) [corrected]. *Am J Ind Med* 1996;29(6):602–608
 - 12 Pellegrini VD Jr. Osteoarthritis at the base of the thumb. *Orthop Clin North Am* 1992;23(1):83–102
 - 13 Armstrong AL, Hunter JB, Davis TR. The prevalence of degenerative arthritis of the base of the thumb in post-menopausal women. *J Hand Surg [Br]* 1994;19(3):340–341
 - 14 Marks M, Schoones JW, Kolling C, Herren DB, Goldhahn J, Vliet Vlieland TP. Outcome measures and their measurement properties for trapeziometacarpal osteoarthritis: a systematic literature review. *J Hand Surg Eur Vol* 2013;38(8):822–838
 - 15 Marks M, Audigé L, Herren DB, Schindele S, Nelissen RG, Vliet Vlieland TP. Measurement properties of the German Michigan Hand Outcomes Questionnaire in patients with trapeziometacarpal osteoarthritis. *Arthritis Care Res (Hoboken)* 2014;66(2):245–252
 - 16 Chung KC, Pillsbury MS, Walters MR, Hayward RA. Reliability and validity testing of the Michigan Hand Outcomes Questionnaire. *J Hand Surg Am* 1998;23(4):575–587
 - 17 Bellamy N, Campbell J, Haraoui B, et al. Dimensionality and clinical importance of pain and disability in hand osteoarthritis: Development of the Australian/Canadian (AUSCAN) Osteoarthritis Hand Index. *Osteoarthritis Cartilage* 2002;10(11):855–862
 - 18 Ware JE, Gandek B. The SF-36 Health Survey: development and use in mental health research and the IQOLA Project. *Int J Ment Health* 1994;23(2):49–73
 - 19 Levine DW, Simmons BP, Koris MJ, et al. A self-administered questionnaire for the assessment of severity of symptoms and functional status in carpal tunnel syndrome. *J Bone Joint Surg Am* 1993;75(11):1585–1592
 - 20 MacDermid JC, Turgeon T, Richards RS, Beadle M, Roth JH. Patient rating of wrist pain and disability: a reliable and valid measurement tool. *J Orthop Trauma* 1998;12(8):577–586
 - 21 Kremer E, Atkinson JH Jr. Pain measurement: construct validity of the affective dimension of the McGill Pain Questionnaire with chronic benign pain patients. *Pain* 1981;11(1):93–100
 - 22 Eaton RG, Littler JW. Ligament reconstruction for the painful thumb carpometacarpal joint. *J Bone Joint Surg Am* 1973;55(8):1655–1666
 - 23 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. Doi: 10.1186/1471-2288-10-22
 - 24 Model Z, Liu AY, Kang L, Wolfe SW, Burket JC, Lee SK. Evaluation of Physical Examination Tests for Thumb Basal Joint Osteoarthritis. *Hand (NY)* 2016;11(1):108–112
 - 25 Price DD, McGrath PA, Rafii A, Buckingham B. The validation of visual analogue scales as ratio scale measures for chronic and experimental pain. *Pain* 1983;17(1):45–56
 - 26 McGraw KO, Wong SP. Forming inferences about some intraclass correlation coefficients. *Psychol Methods* 1996;1:30. Doi: 10.1037/1082-989X.1.1.30
 - 27 Rice ME, Harris GT. Comparing effect sizes in follow-up studies: ROC Area, Cohen's d, and r. *Law Hum Behav* 2005;29(5):615–620
 - 28 Richards JM Jr, Hemstreet MP. Measures of life quality, role performance, and functional status in asthma research. *Am J Respir Crit Care Med* 1994;149(2 Pt 2):S31–S39, discussion S40–S43
 - 29 Boyce MB, Browne JP, Greenhalgh J. The experiences of professionals with using information from patient-reported outcome measures to improve the quality of healthcare: a systematic review of qualitative research. *BMJ Qual Saf* 2014;23(6):508–518
 - 30 Andrawis JP, Chenok KE, Bozic KJ. Health policy implications of outcomes measurement in orthopaedics. *Clin Orthop Relat Res* 2013;471(11):3475–3481
 - 31 Florack TM, Miller RJ, Pellegrini VD, Burton RI, Dunn MG. The prevalence of carpal tunnel syndrome in patients with basal joint arthritis of the thumb. *J Hand Surg Am* 1992;17(4):624–630
 - 32 Amadio PC, Millender LH, Smith RJ. Silicone spacer or tendon spacer for trapezium resection arthroplasty—comparison of results. *J Hand Surg Am* 1982;7(3):237–244
 - 33 Lins RE, Gelberman RH, McKeown L, Katz JN, Kadiyala RK. Basal joint arthritis: trapeziectomy with ligament reconstruction and tendon interposition arthroplasty. *J Hand Surg Am* 1996;21(2):202–209
 - 34 Lovell ME, Nuttall D, Trail IA, Stilwell J, Stanley JK. A patient-reported comparison of trapeziectomy with Swanson Silastic implant or sling ligament reconstruction. *J Hand Surg [Br]* 1999;24(4):453–455
 - 35 Tomaino MM, Pellegrini VD Jr, Burton RI. Arthroplasty of the basal joint of the thumb. Long-term follow-up after ligament reconstruction with tendon interposition. *J Bone Joint Surg Am* 1995;77(3):346–355
 - 36 Massy-Westropp N, Krishnan J, Ahern M. Comparing the AUSCAN Osteoarthritis Hand Index, Michigan Hand Outcomes Questionnaire, and Sequential Occupational Dexterity Assessment for patients with rheumatoid arthritis. *J Rheumatol* 2004;31(10):1996–2001
 - 37 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;33(1):9–17
 - 38 Citron N, Hulme CE, Wardle N. A self-administered questionnaire for basal osteoarthritis of the thumb. *J Hand Surg Eur Vol* 2007;32(5):524–528
 - 39 Greenslade JR, Mehta RL, Belward P, Warwick DJ. Dash and Boston questionnaire assessment of carpal tunnel syndrome outcome: what is the responsiveness of an outcome questionnaire? *J Hand Surg [Br]* 2004;29(2):159–164
 - 40 De Smet L, Sioen W, Spaepen D, van Ransbeeck H. Treatment of basal joint arthritis of the thumb: trapeziectomy with or without tendon interposition/ligament reconstruction. *Hand Surg* 2004;9(1):5–9
 - 41 Fowler A, Swindells MG, Burke FD. Intra-articular corticosteroid injections to manage trapeziometacarpal osteoarthritis—a systematic review. *Hand (NY)* 2015;10(4):583–592
 - 42 Whitehouse JD, Friedman ND, Kirkland KB, Richardson WJ, Sexton DJ. The impact of surgical-site infections following orthopedic surgery at a community hospital and a university hospital: adverse quality of life, excess length of stay, and extra cost. *Infect Control Hosp Epidemiol* 2002;23(4):183–189
 - 43 Belatti DA, Phisitkul P. Trends in orthopedics: an analysis of Medicare claims, 2000–2010. *Orthopedics* 2013;36(3):e366–e372