The Spanish Version of the QuickDASH. Reliability, Measurement Error, Construct Validity and Responsiveness for Outcomes Assessment in Carpal Tunnel Syndrome

La versión en español del QuickDASH: Fiabilidad, error de medición, validez de constructo y sensibilidad para la evaluación de resultados en el Síndrome del Túnel Carpiano.

Roberto S. Rosales[1]   Hugo Sambad-Rodriguez[1]   Carmen Menaya-Fernandez[1]
Yolanda Martin-Hidalgo[2]   Luis Reboso-Morales[2]

[1] Unit for Hand and Microsurgery, GECOT, La Laguna, Santa Cruz de Tenerife, Spain
[2] Department of Orthopaedics, University Hospital of La Candelaria, Santa Cruz de Tenerife, Spain

Address for correspondence   Roberto S. Rosales, MD, MSCR, PhD, Unit for Hand and Microsurgery, GECOT, La Laguna, Santa Cruz de Tenerife, Spain (e-mail: titorosales@telefonica.net).

Abstract

Objective   This paper's purpose was to assess the reliability, measurement error, construct validity, and responsiveness of the Spanish version of the QuickDASH for outcomes assessment in Carpal Tunnel Syndrome (CTS).

Methods   A total population of 40 patients was diagnosed with CTS. The Spanish versions of the QuickDASH, CTS-AL, and EQ-5D were completed by the patients 1 week, the day before surgery, and 3 months after the surgery. Internal consistency was analyzed by Cronbach α coefficient. The intraclass correlation ($ICC_{2,1}$), Lin's coefficient (CCC), and Bland and Altman Limit of Agreement (LoA) were used for the absolute agreement assessment, and the non-parametric Passing-Bablock (P-B) were used to assess constant and proportional systematic bias between the measurements. Cross-sectional precision was analyzed with the Standard Error of the Measurement (SEM). Longitudinal precision for the test-retest reliability coefficient was analyzed with the Standard Error of the Measurement difference (SEMdiff) and the Minimal Detectable Change at 90% (MDC90) and 95% (MDC95) confidence levels. For assessing construct validity, we hypothesized that the QuickDASH would have a strong positive correlation with the CTS-AL and a moderate negative correlation with the EQ-5D Index, using a Pearson correlation coefficient (r) with a level of significance of 0.05.

Keywords
► reliability
► validity
► construction validity
► QuickDASH
► Carpal Tunnel Syndrome

Results Cronbach coefficient was 0.912. $ICC_{2,1}$ and CCC showed a high absolute agreement (0.868 and 0.738 respectively). The P-B regression showed no significant constant and proportional differences between the two administrations of QuickDASH. The QuickDASH showed a responsiveness (ES = 2.1; SRM = 1.97) lower than the CTS-AL (ES = 3.53; SRM = 3.50), and higher than the EQ-5D Index (ES =0.78; SRM =0.83).

Conclusion   The Spanish (Spain) version of the QuickDASH instrument presented a good level of reliability, construct validity, and responsiveness for outcomes assessment in CTS.

Resumen

Objetivo El propósito de este artículo fue evaluar la confiabilidad, el error de medición, la validez de construcción y la respuesta de la versión española del QuickDASH para la evaluación de resultados en el síndrome del túnel carpiano (STC).

Métodos   Una población total de 40 pacientes con diagnóstico de STC completó las versiones en español del QuickDASH, CTS-AL y EQ-5D una semana antes de la cirugía, el día previo a la cirugía y tres meses después de la intervención. La consistencia interna fue analizada mediante el coeficiente α de Cronbach. Para evaluar la concordancia absoluta se utilizaron el coeficiente de correlación intraclase ($ICC_{2,1}$), el coeficiente de correlación de Lin (CCC) y el límite de acuerdo de Bland y Altman (LoA), mientras que el método no paramétrico de Passing-Bablok (P-B) se empleó para evaluar el sesgo sistemático constante y/o proporcional entre las mediciones. La precisión transversal se analizó con el error estándar de medición (SEM). La precisión longitudinal para el coeficiente de fiabilidad test-retest se analizó con el error estándar de la diferencia de medición (SEM_{diff}) y el cambio mínimo detectable al 90 % (MDC_{90}) y al 95 % (MDC_{95}) de nivel de confianza. Para evaluar la validez de construcción, se formuló la hipótesis de que el QuickDASH tendría una fuerte correlación positiva con el CTS-AL y una correlación negativa moderada con el índice EQ-5D, utilizando el coeficiente de correlación de Pearson (r) con un nivel de significancia de 0.05.

Palabras clave
► fiabilidad
► validez
► validez de construcción
► QuickDASH
► Síndrome del Túnel Carpiano

Resultados   El coeficiente de Cronbach fue de 0.912 $ICC_{2,1}$ y CCC mostraron una concordancia absoluta alta (0.868 y 0.738 respectivamente). La regresión P-B no mostró diferencias constantes y proporcionales significativas entre las dos administraciones del QuickDASH. El QuickDASH mostró una respuesta (ES = 2.1; SRM = 1.97) inferior al CTS-AL (ES = 3.53; SRM = 3.50) y superior al índice EQ-5D (ES = 0.78; SRM = 0.83).

Conclusión   La versión española (España) del instrumento QuickDASH presentó buen nivel de confiabilidad, validez de construcción y respuesta para la evaluación de resultados en STC.

Introduction

The use of health outcomes instruments, called patients reported outcomes (PRO) has grown in hand surgery clinical research. One of the most used PRO instruments has been the DASH (Disability of the Arm, Hand, and Shoulder). In 2005, Beaton et al[1] by comparing three item reduction approaches (concept-retention method, equidiscriminative item-total correlation, and Item Response theory or Rasch modeling) developed a shorter version to reduce respondent burden while the PRO instruments kept the psychometric properties. QuickDASH was the shortened version of the DASH Outcome Measure. It was constituted by 11 items, instead of 30 items, to measure physical function and symptoms in individuals with one or more musculoskeletal disorders of the upper. Considering that a Spanish version of the DASH was already available with a good level of reliability, validity, and responsiveness,[2–4] a new Spanish version of the shorter version, QuickDASH, was released and approved by the Institute for Work and Health (IWH), Ontario, Canada (https://dash.iwh.on.ca/sites/dash/public/translations/QuickDASH_Spanish_Spain_2018.pd) However, no new evidence has been reported about the psychometric properties of the Spanish version of the QuickDASH instrument.

This paper's purpose was to assess the reliability, measurement error, construct validity, and responsiveness of the

Spanish version of the QuickDASH for outcomes assessment in Carpal Tunnel Syndrome (CTS).

Methods

Study Population

The study population consisted of 40 consecutive patients (27 women), mean age 54.5 (SD 11.2) years (►Table 1) with the diagnosis of CTS based on clinical and neurophysiological criteria, recruited from the waiting list for CTS surgery in the National Health System in Tenerife, Spain. All procedures performed in this study involving human participants were by with the ethical standards of the institutional national research committee of the University Hospital of La Candelaria, School of Medicine, University of La Laguna, and with the 1964 Helsinki Declaration and its later amendments or comparable ethical standards. The ethics committee reviewed and approved this study (Protocol PI-11/16). Written informed consent was obtained from all individual participants included in the study.

Eligibility Criteria

Inclusion criteria were: (1) numbness or tingling with or without pain in at least 2 of the digits of the median nerve distribution,[5,6] (2) increased symptoms with carpal tunnel provocation tests (Phaleń test, and/or reverse Phalen),[6] (3) symptoms for over two months,[5] (4) failure in conservative treatment,[5] and nerve conduction test showing median

Table 1 Demographic data

	Total (N = 40)
Age	
Mean (SD)	54.48 (11.18)
Median (Q1, Q3)	50.5 (47.5, 59.5)
CTS-AL_pre	
Mean (SD)	3.76 (0.68)
Median (Q1, Q3)	3.9 (3.4, 4.2)
QuickDASH1	
Mean (SD)	63.88 (19.46)
Median (Q1, Q3)	65.9 (53.4, 77.3)
Eq. 5D_pre	
Mean (SD)	0.56 (0.31)
Median (Q1, Q3)	0.7 (0.2, 0.8)
Gender	
Male	13 (32.5%)
Female	27 (67.5%)
Affected hand	
Right	26 (65.0%)
Left	14 (35.0%)

Abbreviations: Q1, First quartile; Q3, third quartile; SD, Standard Deviation.

neuropathy at the wrist (distal motor latency > 4.5 milliseconds, wrist-digit sensory latency > 3.5 milliseconds, or sensory conduction velocity at the carpal tunnel segment < 40 m/second.[7,8] Exclusion criteria were: clinical or electrophysiological signs of proximal nerve compression, diabetes or any metabolic disease, rheumatoid arthritis or other general inflammatory diseases.[5,6,9]

Clinical Design

We conducted an observational study with a cross-sectional design for the Standard Error of the Measurement (SRM) and construct validity analyses and a classic cohort design for test-retest reliability and responsiveness assessment which adhered to the STROBE guidelines and was the clinical design used in this research.[10]

Outcomes Instruments and Measures

The standard Spanish versions of the QuickDASH ((www.iwh.on.ca), the Carpal Tunnel Syndrome 6 items de Atroshi-Leryn (CTS-AL)[11] and Euroquol 5D Index (EQ-5D)[12] questionnaires were completed by the patients 1 week before surgery.

The QuickDASH is the shorter version of the DASH PRO instruments developed for measuring "upper extremities disability." It is constituted by 11 items, and it is scored from 0 (best =lowest disability) to 100 (worst = highest disability). At least 10 of the 11 items must be completed for a score to be calculated. The assigned values for all completed responses are simply summed and averaged, producing a score out of five. This value is then transformed to a score out of 100 by subtracting one and multiplying by 25. This transformation is done to make the score easier to compare with other measures scaled on a 0–100 scale. In this paper, it was used the calculation service tools developed by the Institute for Work and Health to obtain the QuickDASH scores. http://www.orthopaedicscore.com/scorepages/disabilities_of_arm_shoulder_hand_score_quickdash.html.[13]

The CTS-AL measured symptoms severity related to CTS. It is constituted of 6 items. Five of the 6 items in the CTS-AL have similar item text as the corresponding items in the 11-item symptom severity scale (Standard CTS-SS questionnaire)[14] and the remaining item (a result of the merger of 2 symptom severity scale items) has text from the 2 items. The CTS-AL has, however, a completely different and improved layout. The scoring is like that for the 11-item symptom severity score; for each patient, the item responses are scored from 1 (best) to 5 (worst) and then averaged for the 6 items to yield a CTS-AL score (only 1 missing item response is allowed).

The EuroQol (EQ-5D) generic health index comprises a five-part questionnaire and a visual analog self-rating scale. In this paper, it was used he EQ-5D as a health-related Quality of Life index with five dimensions: mobility (MO), self-care (SC), usual activities (UA), pain/discomfort (PD), and anxiety/depression (AD). The possible range for each of the dimension variables is 1 to 3, where 1 = no problems, 2 = moderate problems, and 3 = extreme problems. Once the analytical dataset had appropriately named dimension

variables, the EQ-5D Index was generated using the EQ-5D TM Scoring algorithm for Excel Excel http://www.ahrq.gov/professionals/clinicians-providers/resources/rice/EQ5Dscore.html. The EQ-D- Index scores ranged from 0 (death) to 1(best health). Some patients with severe long-standing diseases may have health states which attracted utility values below zero, i.e., from a societal perspective they were regarded as being in states "worse than death"[15]. The index score will not be calculated when responses are missing for one or more of the dimensions. No missing items from the three PRO instruments were observed in this study.

For assessing test-retest reliability a second self-administration of the QuickDASH was done 1 week after the first administration and answered a simple questionnaire consisting of questions regarding changes in their upper extremities' health status during the preceding week. Finally, the QuickDASH was self-administered to the sample population 3 months after open carpal tunnel release for responsiveness analysis.

The measurement property "construct validity" included three aspects: structural validity, hypotheses testing, and cross-cultural validity. Hypothesis testing was used for analyzing the construct validity of the QuickDASH. Construct hypotheses testing validity is the degree to which the scores of a PRO instrument are consistent with hypotheses based on the assumption that the PRO instrument validly measures the construct to be measured.[16–18] For assessing construct validity it was hypothesized a priori that the QuickDASH would have a moderate to strong positive correlation with the CTS-AL and a low to moderate negative correlation with the EQ-5D Index.

Responsiveness is the ability of a PRO instrument to detect change over time in the construct to be measured and it is related to longitudinal validity or change scores over time. For that purpose, an a priori hypothesis was established: the QuickDASH, as an upper extremities-specific instrument would have responsiveness lower than the CTS-AL (disease-specific PRO instrument), and higher than a generic instrument as the EQ-5DIndex.

Data Analysis

Reliability
Internal consistency or the degree to the QuickDASH measured a single concept, was assessed with the Cronbach α coefficient ($\alpha > 0.7$ indicated a good internal consistency). Reproducibility or test-retest reliability was analyzed by the Intraclass correlation coefficient (two-way random effect model and absolute agreement definition ($ICC_{2,1}$),[18] Lińs Concordance Correlation Coefficient (CCC), Bland and Altman Limits of Agreement (LoA) and the non-parametric Passing-Bablok (P-B) orthogonal regression model between the two administrations before surgery, taking 1 week as washout time.

The Bland-Altman limits of agreement (LoA)[19] calculate the difference between both measures for each subject ($di = Y - X$) and it is faced with the mean (($Xi + Yi$)/2) of both measures for everyone. If we assume the normal distribution

of the differences, it is expected that 95% of the differences should be between the limits of the interval as an indicator of a good agreement. The normality of the distribution of the difference was analyzed by the Shapiro-Wilk test; besides, kurtosis and skewness were also analyzed using a level of significance of 0.05.

The Passing–Bablok (P-B) regression line of agreement was used in case we did not need to assume a normal distribution of the differences. The P-B analysis is a non-parametric estimation of the orthogonal regression line between the two methods or measures. The linear equation will be $Y = A + BX + \varepsilon$, in which A is the constant difference between the two measures, B is the proportional difference, and ε is a random variable with a mean equal to zero which represents the random non-systematic error between the two measures. When $A = 0$ and $B = 1$ meant that both measures presented the same error, they presented an excellent agreement, and they were comparable and interchangeable. Based on the P-B non-parametric regression we could determine if there were any significant systematic constant and/or proportional differences between two measures.[18]

Measurement Error
Cross-sectional precision was analyzed with the Standard Error of the Measurement (SEM = SD multiplied by the squared root of 1-Cronbach coefficient). Longitudinal precision for the test-retest reliability coefficient was analyzed with the Standard Error of the Measurement difference (SEMdiff = SD multiplied by the squared root of 1-ICC multiplied by the squared root of 2) and the Minimal Detectable Change at 90% confidence level (MDC90 = SEM diff multiplied by 1.65) and 95% confidence level (MDC95 = SEM diff multiplied by 1.96).[20]

Hypothesis Construct Validity
The construct validity hypotheses were analyzed with the Pearson correlation coefficient (r), using a level 0.05 for statistical significance. Values between 0.8 and 1.0 indicating a very strong relationship, between 0.6 and 0.8 a strong relationship, between 0.4 and 0.6 a moderate relationship, between 0.2 and 0.4 a weak relationship, and less than 0.2 very weak or no relationship.[21]

Responsiveness
The responsiveness of every PRO instrument was evaluated by calculating the effect size (ES = mean change scores /standard deviation (SD) of the baseline scores) and the standardized response mean (SRM = mean change scores / SD of the change). A large SRM or ES indicated high sensitivity to change, and an ES or SRM >0.8 meant an important clinical improvement.[18]

Sample Size
For an expected ICC of 0.85 in the test-retest reliability, a 95% confident Interval (CI) width of 0.20, 2 measures, the expected sample size was 31. For construct validity, a priori sample size calculation for the correlation analysis showed that based on the proposed null hypothesis (Ho = the

correlation is equal to zero), with the 2-sided test, 0.05 significance level, 80% power and expected minimum r of 0.4, a sample size of 37 patients would be needed (Stata 16.1. StataCorp 4905. Lakeway Drive. College Station, Texas 77845 USA)

Results

Reliability

Internal consistency analysis showed a Cronbach α of 0.912 with an average interitem covariance of 0.522. Absolute test re-test agreement analysis demonstrated an $ICC_{2,1}$ of 0.868 (95% C.I.: 0.750 to 0.930) ($p < 0.001$) and a CCC of 0.738 (95% C.I.: 0.597 to 0.877) ($p < 0.001$). Bland- Altman analysis showed that the estimated absolute difference bias between both administrations of the QuickDASH was 5.47 units (SD 13.39) with a LoA of -20.767 to 31.707 being 95% of the cases included in the interval (2 cases over the limit which meant a 5% and no one case under the limit (►**Fig. 1**). However, the assumption of normality of the difference's distribution was not confirmed (Shapiro-Wil test $p = 0.005$, Skewness test $p = 0.0019$, Kurtosis test $p = 0.0029$).

P-B regression showed a median difference between both administrations of the QuickDASH scores in the test re-test

reliability of 2.3 units with no significant constant and proportional differences (►**Table 2**) (►**Fig. 2**).

Measurement Error

Cross-sectional precision analysis demonstrated an SEM of 5.785. Longitudinal precision showed a SE_{diff} of 10.0001 with an $MDC_{95} = 19.6$ and an $MDC_{90} = 16.5$ (►**Table 3**).

Construct Validity

The QuickDASH presented a positive strong correlation with the CTS-AL ($r = 0.635$) and negative a moderate correlation with the EQ-5D Index ($r = -0.492$), being significant ($p < 0.001$) (Supplementary material).

Responsiveness

The QuickDASH showed a responsiveness (ES = 2.1; SRM = 1.97) lower than the CTS-AL (ES = 3.53; SRM = 3.50), and higher than the EQ-5D Index (ES = 0.78; SRM = 0.83) (►**Table 4**).

Discussion

The results have demonstrated that the QuickDASH PRO instrument presented a good level of internal consistency
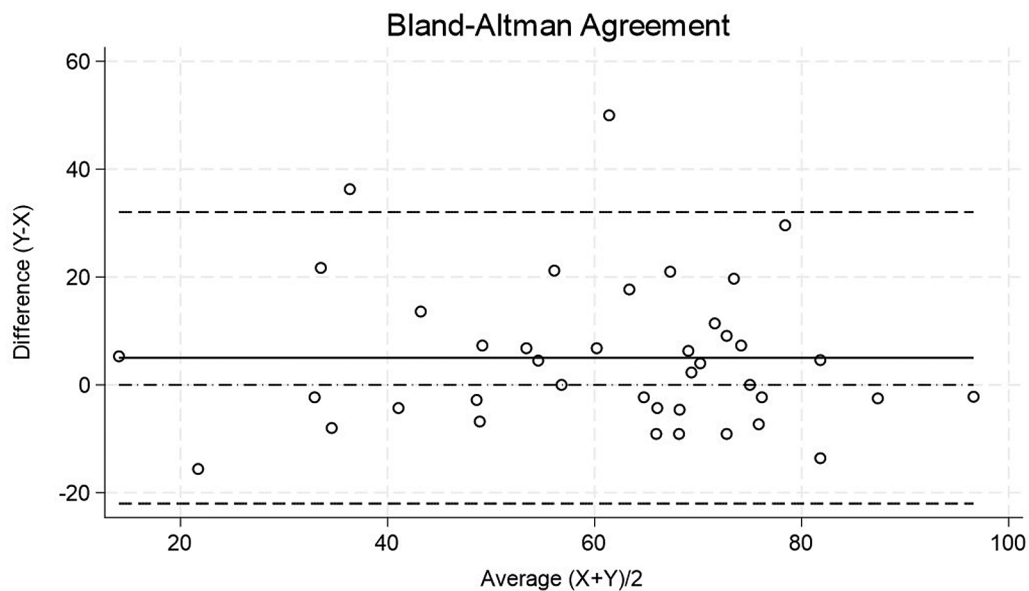

Fig. 1 Bland-Altman Limits of Agreement. The Bland-Altman limits of agreement (LoA) calculate the difference between both measures for each subject (di = Y – X) and it is faced with the average ((Xi + Yi)/2) of both measures for everyone. It is shown that 95% of the cases were included in the interval (only 2 cases over the limit which meant 5% and no one case under the limit).

Table 2 Test – retest Passig-Bablok agreement. Descriptive statistics (listwise)

Variable	Valid	Miss	Obs	Median	Mean	Minimum	Maximum	SD
Y: QDash1	40	0	40	65.9	63.88	13.6	95.5	19.45837
X: QDash2	40	0	40	63.6	58.41	11.4	97.7	19.73363
Y-X	40	0	40	2.3	5.47	-15.9	50	13.38622
100*(Y-X)/X	40	0	40	4.58%	16.0%	-30.5%	199.5%	43.0%

Abbreviations: QDash1, QuickDash first administration; QDash2, QuickDash second administration 1 week apart; SD, Standard Deviation.
$100^*(Y-X)/X$ = percentage of the differences taking the QDash2 as reference.

Passing Bablok Regression line (Y = A + Bx)

A = 2.3 (95% CI: -9.71 to 16.06)
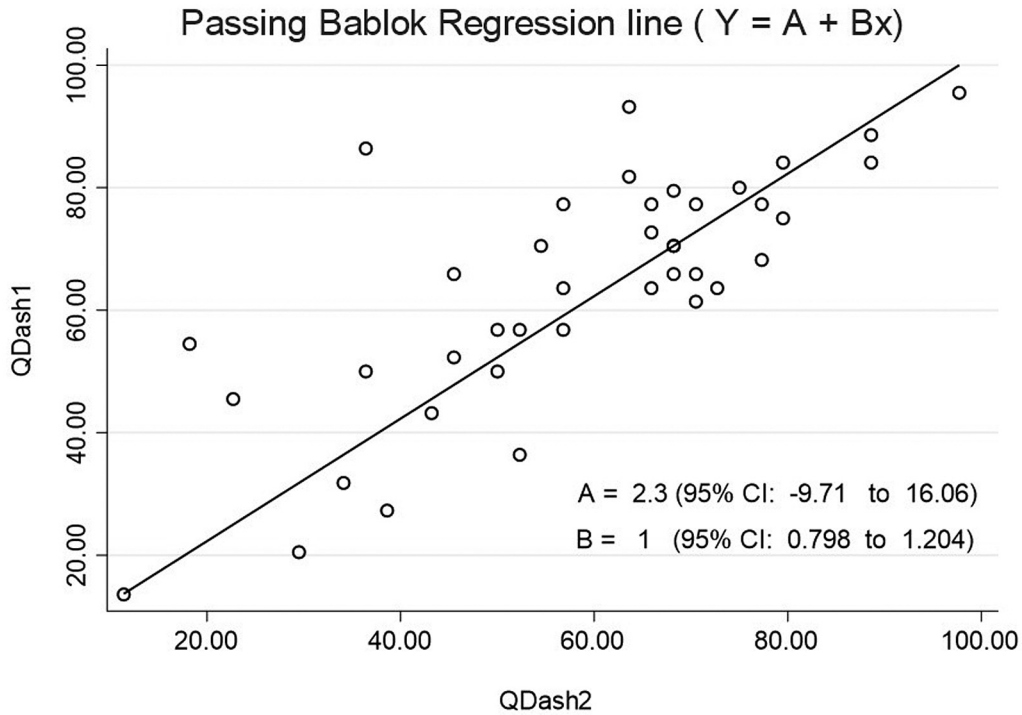
B = 1 (95% CI: 0.798 to 1.204)

Fig. 2 Passing Bablok Regression Line. A is the constant difference between the two measures, and B is the proportional difference. If A = 0 and B = 1 meant that both measures presented the same error, and they presented an excellent agreement. Observe that the null hypotheses (A = 0 and B = 1) were included in the 95% confidence interval and the P-B showed no significant constant and proportional differences.

Table 3 Measurement error of the QuickDASH

A) Cross sectional precision				
Cronbach α	SD	SEM	Cross Sectional Precision. 95% CI	Cross Sectional precision. 90% C.I.
0.912	19.458	5.785	-/+ 11.339	-/+ 9.546
B) Longitudinal precision.				
$ICC_{2,1}$	SD	SE_{diff}	MDC_{95}	MDC_{90}
0. 868	19.4584	10.000	19.600	16.500

Abbreviations: CI, Confident Interval; ICC, Intraclass correlation coefficient; MDC, smallest detectable change (SDC) define by the COSMIN; MDC_{90}, Minimal Detectable Change at 90% confidence level; MDC_{95}, Minimal Detectable Change at 95% confidence level; SD, Standard Deviation; SE_{diff}, Standard Error of the Measurement difference; SEM, Standard Error of the Measurement.

and a high level of test-retest reliability with an excellent level of absolute agreement coefficients ($ICC_{2,1}$, CCC) without significant constant or proportional systematic differences between the scores of two administrations, 1 week apart. Results of validity and responsiveness analyses were coincidental with the construct hypotheses formulated a priori as proof of good cross-sectional and longitudinal construct validity.

Different tools have been developed for assessing the quality of a PRO instrument. The work of the original team, from the Medical Outcomes Trust, that developed the classic generic instruments, the SF-36 and SF-12, was the basis for two guidelines: "Evaluating the Measurement of Patient-Reported Outcomes" (EMPRO),[22] and "Consensus-based standards for the selection of health measurement instruments" (COSMIN).[17,23] Based on the COSMIN, three main

quality domains could be distinguished to assess a PRO Instrument: reliability, validity, and responsiveness. Each domain may include different measurement properties.[18] Reliability included 3 measurement properties: internal consistency, test-retest reliability, and measurement error. Cronbach's α coefficient is an estimate of internal consistency and depend on the number of items in a scale and their magnitude of intercorrelation and values >0.7 indicate good internal consistency. In this research we observed a Cronbach́s α of 0.912 very close to that one reported by Gabel,[24] Alnahdi,[25] and Claro da Silva[26]; and slightly lower than the value of 0.94 reported by Hammond et al.[27] (British version), Beaton et al.[1] in the development of the QuickDASH and that one reported by Schønnemann et al.[28] (α = 0.96 Danish version) [2016]. Absolute agreement in test-retest reliability analysis demonstrated an excellent level of agreement with

Table 4 Responsiveness and longitudinal construct validity of the QuickDASH

PRO Instrument	Pre-op	Post-op	Change pre-post	Responsiveness	
	Mean (SD)	Mean (SD)	Mean (SD)	ES	SRM
CTS_AL	3.76 (0.68)	1.38 (0.44)	2.38 (0.68)	3.53	3.50
QuickDASH	63.88 (19.46)	23.01 (13.19)	40.87 (20.75)	2.10	1.97
Eq. 5D Index	0.56 (0.31)	0.80 (0.19)	0.24 (0.29)	0.78	0.83

Abbreviations: CTS_AL, Carpal Tunnel Syndrome 6 items de Atroshi- Leryn scale; Eq. 5D, EuroQol scale; ES, Effect Size; Post-op, Post operation; Pre-op, Pre operation; PRO Instrument, Patients Reported Outcomes Instrument; QuickDASH, shorter version of the DASH PRO instrument; SD, Standard Deviation; SRM, Standardized Response Mean.

an $ICC_{2,1}$ of 0.85 and CCC of 0 0.74 very similar to the Brazilian QuickDASH ($ICC_{2,1} = 0.81$)[26] and the Korean version (ICC $_{2,1} = 0.83$)[29] (►**Table 5**). Measurement error is another psychometric property included in the domain of reliability by COSMIN. SEM or cross-sectional precision gives us information about the variation around an observed score. The true score exists within this range. In this research, the SEM of 5.78 was slightly higher compared with that one reported by Beaton et al in the description of the longer version DASH (SEM = 4.4).[1] For longitudinal precision of the measurement we calculated the SE_{diff}, MDC_{90}, and MDC95. The MDC is the minimum change score required before an individual can confidently be considered to have changed by more than day-to-day variability.[30] In this research, our result at the level of $MDC_{90} = 16.5$ was very similar to the range of 11.0–17.2 reported in a Systematic review of the quality of measurement properties of the QuickDASH by Kennedy et al.[31] However, our MDC_{95} 19.6 was lower than that one published by Mintken et al.[32] When we compare MDC we should take in account that this absolute reliability index depends on the study population, washout interval, time point during the follow-up when the test-retest analysis was done, and the SD of the data.[3] Important issue in the discussion of test-retest reliability is that the mean difference score between the two administrations of the QuickDASH 1 week apart (5.47) (►**Table 2**) was lower than the MDC_{95} as a proof that that change should be considered as a change by more than day-to-day variability.

The COSMIN established that the appropriate statistics for assessing measurement error are SEM, LoA, and the smallest detectable change (SDC) or MDC. Changes within the LoA or smaller than the MDC_{95} are likely to be due to measurement error and changes outside the LoA or larger than the SDC should be considered as real change. In this research, we included the LoA as an appropriate statistic for test-retest agreement because Bland and Altman[33] developed the LoA as an alternative analysis to the ICC; but the assumption of normality of the differences between the two measures was violated (Shapiro Wilks $p = 0.005$) in our sample. Consequently, LoA results should not be considered.

Rosales & Atroshi[18] introduced in the hand surgery field the use of the non-parametric P-B analysis to determine constant and/or proportional systematic errors between two measures with the advantage that the assumption of nor-

mality was not necessary. Our result demonstrated that there was no constant (A = 2.3; 95% CI: -9.71 to 16.06) or proportional (B = 1; 95% CI: 0.80 to 1.20) systematic differences between the two administrations of QuickDASH, 1 week apart.

Construct validity analysis confirmed the a priori hypothesis as proof of good validity of the QuickDASH for measuring the proposed construct by a cross-sectional design (correlations between QuickDASH, CTS-AL, and EQ-5D) and by a longitudinal design or responsiveness analysis (comparison of ES and SRM of the QuickDASH, CTS-AL, and EQ-5D from baseline to 3 months after surgery).

Strong point of this study is that based on COSMIN check list,[17,23] the reliability, validity and responsiveness of this paper met most of the design requirements as standards for good methodological quality of a PRO instrument: description of how the missing items were handled, percentage of missing values, sample size calculation, at least two independent measurements with appropriated and stable wash out interval of 1 week in which conditions were similar for both measurements in the reliability analysis, the hypotheses testing regarding correlations and responsiveness were formulated a priori before the data collections, the expected directions of the correlations and effect sizes were included in the hypothesis, the expected absolute or relative magnitude of correlations effects sizes were include in the hypothesis, there was an adequate description of comparator instruments and their properties, and finally adequate clinical design and statistical method for the hypotheses to be tested.

The limitation of this study is the lack of interpretability analysis which can give information about MICD (minimal important clinical difference) or the degree to which one can assign qualitative clinical meaning to an instrument's quantitative scores or change in scores.[17,18,23] Further studies regarding MICD are recommended to complete the analysis of the measurement's properties of the Spanish QuickDASH PRO instrument.

In conclusion, this study has demonstrated that the Spanish (Spain) version of the QuickDASH instrument has good reliability, construct validity, and responsiveness for outcomes assessment in CTS.

Conflict of Interests
The authors declare that there is no conflict of interest.

Table 5 Review quickdash translations from 2011 to 2023

Translated DASH	Study author, year	Study population	Testing reported in study			Responsiveness
			Reliability: Internal Consistency (Cronbachs Alpha)	Reliability: Test-retest (ICC unless stated otherwise)	Validity	
Arabic	Alnahdi 2021[25]	Participants with upper extremity musculoskeletal disorders	QuickDASH 0.90	ICC (2,1) = 0.91	Construct GAF r = −0.53 NRPS r = 0.52 RAND 36: Physical functioning r = −0.77 Emotional well-being r = −0.47 Pain r = −0.75	
Brazilian	Claro da Silva 2020[26]	Participants with Upper extremity disorders	DASH 0.93 QuickDASH 0.88	ICC = 0.81	Construct DASH r = 0.91 SF-12 Physical component r = −0.55 Mental component r = −0.49	Observed change at third interview SRM = 0.94
Chinese	Cao 2019[33]	Patients suffering from upper limb chronic disorders	QuickDASH 0.81	ICC = 0.90	Construct DASH r = 0.820 VAS r = 0.463 SF-36 subscales Physical function r = −0.630 Role physical r = −0.471 Bodily pain r = −0.563 General health r = −0.414 Vitality r = −0.053 Social function r = −0.178 Role emotional r = −0.010 Mental health r = −0.165	–
Danish	Schønnemann 2016[28]	Patients with distal fracture radius	QuickDASH 0.96	ICC = 0.94	–	–
English (British)	Hammond 2018[27]	Patients with rheumatoid arthritis	DASH 0.98 DASH Work 0.94	ICC (2,1) = 0.95	Concurrent QuickDASH r = 0.61–0.91 DASH r = 0.61–0.99 DASH Work r = 0.53–0.80 DASH Spam r = 0.52–0.78	
Italian	Franchignoni 2011[34]	Patients with upper limb disorders	QuickDASH 0.87	–	–	–
Korean	Hong 2018[29]	Patients with Carpal Tunnel Syndrome	QuickDASH 0.89 ADL 0.89	ICC (2,1) = test for each items ranged 0.64 to	Construct DASH r = 0.975 Pain VAS r = 0.365	Observed change six months after operation SRM = 1.00

(*Continued*)

Table 5 (*Continued*)

Translated DASH	Study author, year	Study population	Testing reported in study			Responsiveness
			Reliability: Internal Consistency (Cronbach's Alpha)	Reliability: Test-retest (ICC unless stated otherwise)	Validity	
			Social Activity 0.70 Symptom 0.72	0.98 and 0.83 for all QuickDASH	Blend Scale of NCS r=−0.074 Grip Power r=−0.309 Pinch power r=−0.327	

Abbreviations: ADL, Activities of daily living; GAF, Global Assessment of Function; ICC, Intraclass correlation coefficient; NRPS, Numeric Pain Rating Scale; SRM, Standardized Response Mean; VAS, Visual Analogue Scale.

References

1 Beaton DE, Wright JG, Katz JNUpper Extremity Collaborative Group. Development of the QuickDASH: comparison of three item-reduction approaches. J Bone Joint Surg Am 2005;87(05): 1038–1046

2 Rosales RS, Delgado EB, Díez de la Lastra-Bosch I. Evaluation of the Spanish version of the DASH and carpal tunnel syndrome health-related quality-of-life instruments: cross-cultural adaptation process and reliability. J Hand Surg Am 2002;27(02):334–343

3 Rosales RS, Diez de la Lastra I, McCabe S, Ortega Martinez JI, Hidalgo YM. The relative responsiveness and construct validity of the Spanish version of the DASH instrument for outcomes assessment in open carpal tunnel release. J Hand Surg Eur Vol 2009;34 (01):72–75

4 Kennedy CA, Beaton DE, Solway S, McConnell S, Bombardier C. The DASH and Quick DASH outcome measure user's manual. 3rd ed. Toronto, Ontario:: Institute for Work & Health;; 2011

5 Atroshi I, Gummesson C, Johnsson R, Sprinchorn A. Symptoms, disability, and quality of life in patients with carpal tunnel syndrome. J Hand Surg Am 1999;24(02):398–404

6 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(02):250–254

7 Atroshi I, Larsson GU, Ornstein E, Hofer M, Johnsson R, Ranstam J. Outcomes of endoscopic surgery compared with open surgery for carpal tunnel syndrome among employed patients: randomised controlled trial. BMJ 2006;332(7556):1473

8 Kimura J. Electrodiagnosis in diseases of nerve and muscle: principles and practice. 2nd ed. Philadelphia:: FA Davis;; 1989

9 Amadio PC, Silverstein MD, Ilstrup DM, Schleck CD, Jensen LM. Outcome assessment for carpal tunnel surgery: the relative responsiveness of generic, arthritis-specific, disease-specific, and physical examination measures. J Hand Surg Am 1996;21 (03):338–346

10 von Elm E, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, Vandenbroucke JPSTROBE Initiative. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. J Clin Epidemiol 2008;61(04):344–349. Doi: 10.1016/j.jclinepi.2007.11.008

11 Rosales RS, Atroshi I. Spanish versions of the 6-item carpal tunnel syndrome symptoms scale (CTS-6) and palmar pain scale. J Hand Surg Eur Vol 2013;38(05):550–551

12 Badia X, Roset M, Montserrat S, Herdman M, Segura A. [The Spanish version of EuroQol: a description and its applications. European Quality of Life scale]. Med Clin (Barc) 1999;112(Suppl 1):79–85

13 Hudak PL, Amadio PC, Bombardier CThe 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(06):602–608

14 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

15 Hurst NP, Kind P, Ruta D, Hunter M, Stubbings A. Measuring health-related quality of life in rheumatoid arthritis: validity, responsiveness and reliability of EuroQol (EQ-5D). Br J Rheumatol 1997;36(05):551–559

16 Rosales RS, Gutiérrez R, Hidalgo Y, Reboso L, Atroshi I. Cómo escribir un artículo original en Cirugía de la Mano. Rev Iberoam Cir Mano 2016;44(01):47–58

17 Mokkink LB, de Vet HCW, Prinsen CAC, et al. COSMIN Risk of Bias checklist for systematic reviews of Patient-Reported Outcome Measures. Qual Life Res 2018;27(05):1171–1179

18 Rosales RS, Atroshi I. The methodological requirements for clinical examination and patient-reported outcomes, and how to test them. J Hand Surg Eur Vol 2020;45(01):12–18

19 Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. Lancet 1986;1 (8476):307–310

20 Kennedy CA, Beaton DE, Solway S, McConnell S, Bombardier C. The DASH and Quick DASH outcome measure user's manual. 3rd ed. Toronto, Ontario:: Institute for Work & Health; 2011

21 Chung MK. Correlation Coefficient. In: Salkin NJ, Ed. Encyclopedia of Measurement and Statistics. London,: Sage Publications,; 2007: 189–201

22 Valderas JM, Kotzeva A, Espallargues M, et al. The impact of measuring patient-reported outcomes in clinical practice: a systematic review of the literature. Qual Life Res 2008;17(02): 179–193

23 Terwee CB, Mokkink LB, Knol DL, Ostelo RW, Bouter LM, de Vet HC. Rating the methodological quality in systematic reviews of studies on measurement properties: a scoring system for the COSMIN checklist. Qual Life Res 2012;21(04):651–657

24 Gabel CP, Michener LA, Melloh M, Burkett B. Modification of the upper limb functional index to a three-point response improves clinimetric properties. J Hand Ther 2010;23(01):41–52

25 Alnahdi AH. Validity and reliability of the Arabic quick disabilities of the arm, Shoulder and Hand (QuickDASH-Arabic). Musculoskelet Sci Pract 2021;53:102372

26 da Silva NC, Chaves TC, Dos Santos JB, et al. Reliability, validity and responsiveness of Brazilian version of QuickDASH. Musculoskelet Sci Pract 2020;48:102163

27 Hammond A, Prior Y, Tyson S. Linguistic validation, validity and reliability of the British English versions of the Disabilities of the Arm, Shoulder and Hand (DASH) questionnaire and QuickDASH in people with rheumatoid arthritis. BMC Musculoskelet Disord 2018;19(01):118

28 Schønnemann JO, Eggers J. Validation of the Danish version of the Quick-Disabilities of Arm, Shoulder and Hand Questionnaire. Dan Med J 2016;63(12):A5306

29 Hong SW, Gong HS, Park JW, Roh YH, Baek GH. Validity, Reliability and Responsiveness of the Korean Version of Quick Disabilities of the Arm, Shoulder, and Hand Questionnaire in Patients with Carpal Tunnel Syndrome. J Korean Med Sci 2018;33(40):e249

30 Stratford PW, Binkley J, Solomon P, Finch E, Gill C, Moreland J. Defining the minimum level of detectable change for the Roland-Morris questionnaire. Phys Ther 1996;76(04):359–365, discussion 366–368

31 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 Quick-DASH: a systematic review. Qual Life Res 2013;22(09):2509–2547

32 Mintken PE, Glynn P, Cleland JA. Psychometric properties of the shortened disabilities of the Arm, Shoulder, and Hand Questionnaire (QuickDASH) and Numeric Pain Rating Scale in patients with shoulder pain. J Shoulder Elbow Surg 2009;18(06):920–926

33 Cao S, Zhou R, Zhou H, et al. Reliability and validity of Simplified Chinese version of Quick Disabilities of the Arm, Shoulder, and Hand (QuickDASH) questionnaire: cross-cultural adaptation and validation. Clin Rheumatol 2019;38(11):3281–3287

34 Franchignoni F, Ferriero G, Giordano A, Sartorio F, Vercelli S, Brigatti E. Psychometric properties of QuickDASH - a classical test theory and Rasch analysis study. Man Ther 2011;16(02): 177–182