MSJ Experimental Translational Clinical

Original Research Paper

Patient-reported outcomes in multiple sclerosis: Validation of the Quality of Life in Neurological Disorders (Neuro-QoL™) short forms

Multiple Sclerosis Journal—Experimental, Translational and Clinical

October–December 2019, 1–11

DOI: 10.1177/2055217319885986

Luis D Medina, Stephanie Torres, Enrique Alvarez, Brooke Valdez and Kavita V Nair

Abstract

Background: Patient-reported outcome (PRO) measures have been shown to be effective for tracking treatment outcomes in multiple sclerosis (MS). However, collecting PROs as part of the clinical standard of care can be time-consuming and examination of their validity for use in an MS sample has been limited.

Objective: To determine the discriminant validity of the Quality of Life in Neurological Disorders (Neuro-QoL™) short forms in a real-world MS clinic population.

Design/Methods: Neuro-QoL is a series of questionnaires for tracking physical function, emotional/cognitive health, and social abilities in clinical populations. Neuro-QoL data from 902 MS patients were analyzed for psychometric properties and factor structure.

Results: Neuro-QoL demonstrated acceptable reliability in the moderate-to-good ranges. Moderate support for convergent validity was observed with other measures of MS quality of life, disease severity, and symptoms. However, results from a confirmatory factor analysis suggested poor model fit for most of the 12 domains tested.

Conclusions: These findings support the utility of some of the Neuro-QoL questionnaires in evaluating MS-related PROs. However, additional research may help abridge and strengthen these measures for use in this population.

Keywords: Multiple sclerosis, patient-reported outcomes, validity, reliability, principal component analysis

Date received: 29 April 2019; accepted: 8 October 2019

Multiple sclerosis (MS) is a chronic autoimmune disease of the central nervous system consisting of both inflammatory and neurodegenerative attributes and characterized by demyelination and axonal degeneration.[1] Due to its complex pathological nature, MS is associated with various clinical symptoms and disabilities, including but not limited to motor impairments, weakness, pain, incontinence, fatigue, sensory difficulties, psychiatric features, and cognitive dysfunction/impairments.[1,2] Given the clinical variability seen in MS, it is crucial to assess quality of life in these patients and how their symptoms affect their day-to-day functioning.

Patient-reported outcome (PRO) measures have been shown to be effective for tracking outcomes in MS. However, they can be time-consuming and undervalued. A challenge of implementing PROs in daily clinical practice and research is that time constraints may limit the ability of busy medical providers and research staff to maintain the use of these measures.[3] Despite the barriers to their implementation, PROs promote patient-centered care in a number of ways: (a) allowing patients to have another means of communication with providers regarding their symptoms; (b) providing information that may not otherwise be communicated, which in turn leads

Correspondence to:
Luis D Medina,
Department of Psychology, University of Houston, 3695 Cullen Boulevard, Room 126, Houston, TX USA.
ldmedina2@uh.edu

Luis D Medina, **Stephanie Torres**, Department of Psychology, University of Houston, USA

Enrique Alvarez, **Brooke Valdez**, Department of Neurology, University of Colorado School of Medicine, USA

Kavita V Nair, Department of Neurology,

University of Colorado School of Medicine, USA
University of Colorado, Skaggs School of Pharmacy and Pharmaceutical Sciences, USA

to clinical action; and (c) providing clinicians with visual quantitative values that may provide further insight on symptom severity.[4] In 2001, the Institute of Medicine called for a shift in patient-centered care, suggesting patients should be allowed to have a voice in their care by considering their personal preferences, values, needs, and lifestyles[5]; PROs have been found to assist with this shift. Patients have endorsed the value of the information gained from completing PROs and are more likely to complete them when provided if they are valued and prioritized as a way to improve their care.[4] Despite this value, translating the quality-of-life data provided by PROs has been a challenge. This appears to be primarily due to a lack of standardization across different PROs as well as the PROs having questionable relevance for certain patient populations, resulting in a lack of generalizability of their outcomes.[6–9] Therefore, it is vital to validate these measures for use in their intended clinical populations.

The progressive nature of MS pervasively impacts patients' physical, social, emotional, and cognitive functioning. This has resulted in the development of numerous MS-specific PRO scales. Modern psychometric methods, such as item response theory, have improved the precision and accuracy of PRO measures, their utility across a variety of chronic disease states, and their ability to be administered in a variety of formats.[10] This prompted the National Institutes of Health Quality of Life in Neurological Disorders (Neuro-QoL™) measurement initiative. Neuro-QoL is a comprehensive system of PRO measures that target neurological disorders. They include item banks and short forms (SFs) for measuring physical, social, and mental domains of health-related quality of life.[11]

The Neuro-QoL is intended to be used in the following neurological disorders: stroke, MS, amyotrophic lateral sclerosis, Parkinson's disease, epilepsy, and muscular dystrophy. Since its release into the public domain in 2012, validation of the SFs of Neuro-QoL in the MS population has been limited. However, a number of validation studies have noted that the Neuro-QoL SFs appear to be valid measures in adults with neurological dysfunction, such as MS[12] and epilepsy,[9] and to have good psychometric properties (e.g., internal consistency, test–retest reliability) for assessing functioning in individuals with neurological disorders.

In the current study we wanted to continue the validation efforts of Neuro-QoL in the MS population and sought to determine the discriminant validity of the Neuro-QoL SF scales in a clinical population of MS patients. In instances when there are high intercorrelations between variables, assessing discriminant validity is necessary for confident interpretation of outcomes.[13] The results of this study will allow clinicians to feel confident that the items within the scales are measuring the target construct.[14] Confirmatory factor analysis (CFA), principal component analysis (PCA), and discriminant analysis allow researchers and clinicians to use the measures that are most efficient and to revise or eliminate measures that are redundant or do not work. This ensures that the scale/measure being used is appropriate for the population of interest.[15]

Methods

Sample

Data were pulled from MS patients seen at the Rocky Mountain Multiple Sclerosis Center at the University of Colorado. Since 2014, as part of their standard of care, patients have been asked to complete a set of PROs annually. A core set of PROs, including the Neuro-QoL SF scales, are captured in one of two ways: (1) an email link containing the PROs that are HIPAA (Health Insurance Portability and Accountability Act) compliant is sent to patients a week ahead of their regularly scheduled clinic visit or (2) during a subsequent clinic visit where patients are provided a tablet on which to complete the PROs . Their responses are then directly fed into a HIPAA-compliant database for analysis. A total of 902 records between 2014 and 2016 were identified for patients who were receiving any type of disease-modifying therapy and were included in the current analyses.

Measures

Patients were diagnosed following suggested guidelines by board-certified neurologists with neuroimmunology training.[16]

Neuro-QoL short forms

The Neuro-QoL SFs are fixed-length questionnaires comprised of items from a larger bank of calibrated items that assess several quality-of-life domains, including physical, mental, and social domains. Patients completed SFs for the following 12 selected domains from the Neuro-QoL Adult Version 1.0 (see Supplemental Table S1): physical function (Upper Extremity Function/fine motor, Lower Extremity Function/mobility), physical symptoms (Sleep Disturbance, Fatigue), emotional health (Anxiety,

Depression, Positive Affect & Well-Being, Emotional & Behavioral Dyscontrol), cognitive health (applied cognition: General Cognitive Concerns, applied cognition: Executive Function, Communication), and social abilities (Ability to Participate in Social Roles & Activities). The format of the questionnaires is similar across SFs. Individuals are generally asked to answer between five and nine questions about how they have been feeling or functioning lately (Communication has five questions, Positive Affect & Well-Being has nine questions, all other SFs have eight questions). Possible responses are all on a five-point scale (e.g., "never" to "always," "without any difficulty" to "unable to do") with a recall period of "In the last seven days." Scoring produces raw scores as well as standardized *T*-scores for comparison with normative and clinical samples. Given that Neuro-QoL *T*-scores function as a conversion of raw scores based on a normative sample, their interpretation requires additional context. For purposes of the current analyses, individual, item-level raw scores were used for a total of 94 Neuro-QoL items across the 12 SFs.

Other PRO measures. Patients completed the Patient-Determined Disease Steps (PDDS),[17] a PRO measure of disability in MS that is both economical and efficient.[18] The PDDS is a measure that allows for evaluating disease progression. Patient motor functioning is rated using a scale ranging from 0–8 (normal to bedridden), which is used to assess disability and mobility in MS. In addition to this, Item 1 from RAND Health Care's 36-Item Short Form Health Survey Instrument (SF-36), a health-related quality of life measure, was used to assess overall health status ("In general, would you say your health is . . . ?") on a 1 through 5 scale, with higher numbers indicating poorer health.[19] Separate items assessing bowel and bladder functioning independently were also administered given that individuals with MS often suffer from such issues. These items assessed functioning on a slider scale, from "0 = not at all" to "100 = severely."

Analysis

We examined the psychometric characteristics of the Neuro-QoL SFs in several ways. CFA was performed on Mplus Version 7.2[20]; all other analyses were performed on SPSS Version 25.[21] Following similar procedures as have been described in the context of Neuro-QoL in epilepsy[9] and MS,[12] the reliability of Neuro-QoL SF scores was assessed using Cronbach's alpha coefficient to examine internal consistency; coefficient values equal to or greater than 0.70 were considered acceptable, suggesting scale items are measuring the same underlying construct.[22] Convergent validity, a component of construct validity, with disease severity was examined using Spearman's rho. The following guidelines were used to interpret magnitude: nominal < 0.30, small $= 0.30$ to 0.49, medium $= 0.50$ to 0.69, and large $= 0.70$ to 1.00.[9] Known-group validity was examined as the extent of association between Neuro-QoL SFs with available measures of similar concepts (i.e., SF-36[23] Item 1, bowel function, bladder function) using analysis of variance (ANOVA). We expected weaker relations between measures of dissimilar constructs and stronger associations between measures of similar or identical ones.

For the purposes of examining the factor structure of the Neuro-QoL SFs in our MS sample, CFA was used. CFA is an inferential method to examine hypothesized a priori models.[24] Using the framework suggested by Neuro-QoL (Figure 1), CFA was conducted to examine the theoretical relationships among our observed and unobserved (latent) variables; in this way, CFA attempts to minimize the difference between the estimated and observed covariance matrices in the data.[25] As illustrated in the framework, Neuro-QoL is theoretically organized into several levels of nested domains. The individual SFs (first level, or 1°) are manifest variables nested within a second level (2°) that consists of proposed latent domains: Function/Health, Symptoms, Emotional Health, Cognitive Health, and Social Abilities. These latent domains are then nested in a model comprised of the Physical, Mental, and Social domains (3°) that help capture overall Quality of Life (4°). For the purpose of the current analyses, only the 1° model was tested such that a separate CFA was carried out for each first order domain. When testing a predetermined model, several indices are used to identify adequate fit of the model to the data. For continuous data, in addition to the X^2 goodness-of-fit index, which is limited due to its sensitivity to sample size, recommended indices include root mean square error of approximation (RMSEA), the Tucker–Lewis Index (TLI), and the Comparative Fit Index (CFI).[25] Recommended cutoffs for these indices were used such that good fit would be indicated by RMSEA < 0.06, TLI ≥ 0.95, and CFI ≥ 0.95.[26] Maximum likelihood estimation was used with a free data format.

To carry out data analytic techniques like CFA, the following assumptions must be met: (a) multivariate

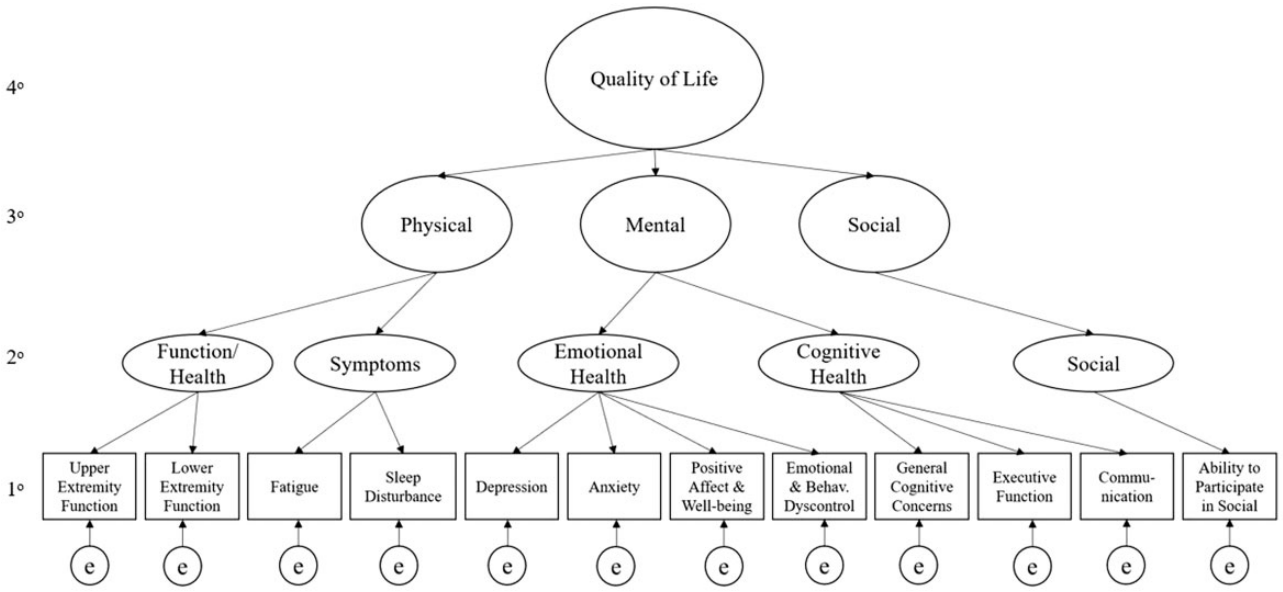

Figure 1. Confirmatory factor analysis model of Neuro-QoL adult domain framework with 12 short forms.

normality within the data must be observed; (b) each factor should comprise at least three variables; (c) the ratio of respondents to variables should be at a minimum 5:1; (d) the correlation (r) between the variables should be 0.30 or greater; (e) if data are missing, it should be in a random pattern; and (f) there should be an absence of multicollinearity and singularity.[27,28]

Results

Sample characteristics

A cohort of 902 patients (mean age 46.8 ± 12.2, mean disease duration 9.1 ± 8.1, 76.1% female) with MS with Neuro-QoL SF data were identified and included in the final analyses. See Table 1 for sample demographics.

Psychometric characteristics

Reliability. Internal consistency was measured by Cronbach's alpha. As shown in Table 2, alphas ranged from 0.85 to 0.97 across the 12 domains measured in our MS sample. These data are shown by MS subgroup in Table 3. Subgroups differed in mean raw scores on SFs related to Social Abilities ($F_{2,828} = 8.22$, $p < 0.001$), Lower Extremity Function ($F_{2,828} = 88.9$, $p < 0.001$), and Upper Extremity Function ($F_{2,828} = 34.5$, $p < 0.001$). Post hoc tests revealed that the relapsing–remitting MS (RRMS) group reported significantly better on Ability to Participate in Social Roles & Activities than the primary progressive MS (PPMS) group ($p = 0.001$); neither groups significantly differed from the secondary progressive MS (SPMS) group. On Lower Extremity Function, the RRMS group reported significantly better functioning than the PPMS and SPMS groups ($p < 0.001$), but the PPMS and SPMS groups did not differ significantly from each other. All three groups differed significantly from each other on self-reported Upper Extremity Function ($p \leq 0.001$) such that the RRMS group reported the best functioning and the PPMS group reported the worst.

Convergent validity. As shown in Table 4, comparison with PDDS, a staging tool for disease severity in MS, as measured with Spearman's rho, demonstrated correlations, in absolute values, ranging from 0.136 to 0.833.

Known-group validity. Known-group analysis with a quality-of-life measure, the SF-36 1-item, as measured with ANOVA, demonstrated significant relationships (all $ps < 0.001$) between Neuro-QoL domains and scores on the SF-36 (Table 5). Worse self-reported quality of life as measured by the SF-36 was associated with worse self-reported functioning on the Neuro-QoL SFs.

Responses on the bowel and bladder scales were not normally distributed; therefore, scores were log transformed and a median split was applied to dichotomize responses to either "low" (<1.0) or "high" (≥ 1.0) scores. Known-group analysis demonstrated significant relationships (all $ps < 0.001$)

Table 1. Sample characteristics: demographic and clinical variables.

Number of patients		902
Mean age in years +/− SD (range)		46.8 ± 12.2 (19−84)
Mean disease duration in years +/− SD (range)		9.1 ± 8.1 (0−45)
Mean PDDS +/− SD (Range)		2.1 ± 2.2 (0−8)
Sex, *n*, %		
Male	216	23.9
Female	686	76.1
Race/ethnicity, *n*, %		
White or Caucasian	667	73.9
Black/African-American	27	3.0
Asian	0	0
Other	36	4.0
American Indian/Alaskan Native	0	0
Hispanic (all races)	0	0
Unknown	172	19.1
Type of MS, *n*, %		
Relapsing–remitting	705	78.2
Secondary progressive	82	9.1
Primary progressive	44	4.9
Unknown	71	7.9

Note: PDDS = Patient-Determined Disease Steps.

Table 2. Descriptive and reliability statistics for Neuro-QoL short forms for the whole sample.

Neuro-QoL short forms	N_{items}	M_{raw} (SD)	M_T (SD)	α	VIF range
Anxiety[a]	8	17.0 (7.3)	50.9 (8.4)	0.94	2.52–4.35
Depression[a]	8	14.0 (6.3)	47.7 (7.2)	0.94	2.06–4.17
Fatigue[a]	8	22.2 (8.5)	49.9 (9.4)	0.96	3.34–7.01
Emotional & Behavioral Dyscontrol[a]	8	16.1 (6.4)	48.5 (9.7)	0.94	2.41–3.93
Sleep Disturbance[a]	8	17.9 (6.2)	51.9 (8.9)	0.85	1.30–2.16
Communication[b]	5	21.9 (3.8)	**	0.88	1.97–2.71
Executive Function[b]	8	34.7 (6.7)	45.2 (10.8)	0.94	2.57–5.14
General Cognitive Concerns[b]	8	28.6 (9.2)	41.5 (9.7)	0.97	3.85–6.46
Positive Affect & Well-Being[b]	9	34.8 (7.1)	53.5 (7.3)	0.95	2.46–9.37
Social Roles & Activities[b]	8	31.7 (7.5)	48.0 (8.2)	0.96	3.24–6.04
Lower Extremity Function[b]	8	34.6 (7.1)	47.9 (10.3)	0.95	2.54–4.86
Upper Extremity Function[b]	8	37.8 (4.2)	46.8 (9.1)	0.92	2.06–3.83

M_{raw} = mean raw scores; M_T = mean T-scores; α = Cronbach's alpha; VIF = variance inflation factor.
[a]Higher score indicates worse functioning.
[b]Higher score indicates better functioning.
**T-scores are not calculated for the Communication scale.

between all Neuro-QoL domains and scores on the bowel and bladder scales such that worse bowel and bladder functioning was associated with worse quality of life (Table 6).

Factor structure. Examination of the data suggested most assumptions were met. Data were verified for normality (Assumption 1) and all components (i.e., domains measured by the SFs) comprised between five and nine items (Assumption 2). Data were collected on a total of 94 questionnaire items across the 12 SFs; our sample of 902 individuals satisfied the recommended minimum 5:1 ratio of respondents to variables (i.e., minimum of 470

Table 3. Descriptive and reliability statistics for Neuro-QoL short forms by subgroup.

	RRMS $n = 705$			SPMS $n = 82$			PPMS $n = 44$			
Neuro-QoL short forms	M_{raw} (SD)	M_T (SD)	α	M_{raw} (SD)	M_T (SD)	α	M_{raw} (SD)	M_T (SD)	α	p
Anxiety[a]	17.3 (7.4)	51.0 (8.5)	0.95	15.6 (6.4)	49.3 (7.7)	0.92	16.5 (7.4)	49.5 (9.2)	0.95	0.132
Depression[a]	14.1 (6.4)	47.7 (7.3)	0.95	13.7 (5.4)	47.6 (6.6)	0.92	15.6 (7.6)	49.1 (8.5)	0.95	0.279
Fatigue[a]	22.3 (8.7)	49.9 (9.6)	0.96	23.0 (7.5)	50.5 (8.2)	0.94	22.5 (7.9)	50.0 (8.9)	0.96	0.788
Emotional & Behavioral Dyscontrol[a]	16.4 (6.6)	48.7 (9.9)	0.95	15.1 (5.0)	47.2 (8.6)	0.91	15.2 (5.7)	47.2 (8.8)	0.92	0.126
Sleep Disturbance[a]	18.0 (6.2)	52.0 (8.9)	0.85	17.5 (6.2)	51.4 (8.7)	0.86	18.1 (6.2)	51.9 (9.5)	0.86	0.776
Communication[b]	22.0 (3.8)	**	0.88	21.9 (3.7)	**	0.83	20.8 (4.5)	**	0.87	0.152
Executive Function[b]	34.9 (6.5)	45.6 (10.6)	0.94	33.8 (7.3)	43.5 (11.4)	0.93	33.0 (9.0)	42.8 (12.1)	0.96	0.096
General Cognitive Concerns[b]	28.4 (9.1)	41.4 (9.7)	0.97	29.0 (9.5)	41.6 (10.2)	0.96	28.6 (9.1)	41.1 (9.6)	0.97	0.863
Positive Affect & Well-Being[b]	34.9 (7.1)	53.6 (7.1)	0.95	33.8 (6.6)	52.5 (6.9)	0.94	33.1 (8.8)	52.3 (9.0)	0.96	0.144
Social Roles & Activities[b]	32.0 (7.4)	48.4 (8.3)	0.97	30.2 (7.2)	45.6 (6.9)	0.95	27.9 (7.3)	44.3 (7.7)	0.92	<0.001
Lower Extremity Function[b]	35.9 (6.2)	49.8 (9.6)	0.94	28.6 (7.7)	39.2 (8.3)	0.92	26.0 (8.0)	36.2 (7.9)	0.92	<0.001
Upper Extremity Function[b]	38.3 (3.7)	47.9 (8.5)	0.92	36.5 (4.7)	42.2 (9.2)	0.89	33.4 (7.4)	38.1 (11.0)	0.94	<0.001

RRMS = relapsing–remitting multiple sclerosis; SPMS = secondary progressive multiple sclerosis; PPMS = primary progressive multiple sclerosis; M_{raw} = mean raw scores; M_T = mean T-scores; α = Cronbach's alpha. ANOVA p-values are provided for raw score comparisons between subgroups.

[a]Higher score indicates worse functioning.

[b]Higher score indicates better functioning.

***T*-scores are not calculated for the Communication scale.

respondents; Assumption 3). Correlations between response items were 0.30 and higher (Assumption 4). It was noted that missing data (Assumption 5) were not missing completely at random (MCAR), as tested by Little's MCAR Test,[29] $p < 0.001$. Given the small number of missing data (a max of 1.2%, or 11 individual responses on any given item), we addressed this violation by computing five multiple data imputations and comparing results across these, as is recommended when MCAR violations are found.[30] Comparison of findings between the original data and the imputed data did not show significant differences. Therefore, the imputed data were used to satisfy the MCAR assumption. Multicollinearity and singularity (Assumption 6) were assessed using the variance inflation factor (VIF) for each SF; VIF values greater than 4 are considered to be of concern, while values greater than 10 are unacceptable.[31] As shown in Table 2, VIF values ranged from acceptable (e.g., Sleep: 1.30–2.16) to of concern (e.g., Fatigue: 3.34–7.01), suggesting possible high collinearity between items within a single SF. However, no VIF values were in the unacceptable range.

Table 4. Spearman's rho correlations for Neuro-QoL short forms raw scores with multiple sclerosis disease severity.

Neuro-QoL short forms	PDDS[a]
Anxiety[a]	0.203
Depression[a]	0.270
Fatigue[a]	0.375
Emotional & Behavioral Dyscontrol[a]	0.136
Sleep Disturbance[a]	0.252
Communication[b]	0.308
Executive Function[b]	0.378
General Cognitive Concerns[b]	0.286
Positive Affect & Well-Being[b]	0.280
Social Roles & Activities[b]	0.547
Lower Extremity Function[b]	0.833
Upper Extremity Function[b]	0.567

Note: PDDS = Patient-Determined Disease Steps. All correlations were significant at the 0.01 level (2-tailed).
[a]Higher score indicates worse functioning.
[b]Higher score indicates better functioning.

The CFA estimation converged normally for all SFs. The X^2 goodness-of-fit test was significant for all SFs (all ps < 0.001). As shown in Table 7, none of the SFs had RMSEA values within the recommended range of <0.06; the range of RMSEA values was 0.079 (lowest, Communication) to 0.202 (highest, Positive Affect & Well-Being). For TLI, only two domains demonstrated acceptable (TLI ≥ 0.95) model fit: Communication (0.976) and General Cognitive Concerns (0.958). These domains also had CFI values in the acceptable (CFI ≥ 0.95) range (Communication: 0.988;

Table 5. Known-group analysis of variance for Neuro-QoL short forms raw scores with SF-36 Item 1.

	SF-36[a]					
Neuro-QoL short forms	1 $n = 70$	2 $n = 288$	3 $n = 346$	4 $n = 184$	5 $n = 13$	Total $n = 901$
Anxiety[a]	12.0 ± 4.8	14.4 ± 6.2	17.3 ± 6.6	21.8 ± 7.6	22.3 ± 9.7	17.0 ± 7.3
Depression[a]	9.8 ± 2.7	11.8 ± 4.9	14.3 ± 5.8	18.2 ± 7.1	19.5 ± 9.5	14.0 ± 6.3
Fatigue[a]	14.8 ± 6.2	18.1 ± 7.3	23.3 ± 7.3	29.0 ± 7.0	28.2 ± 10.6	22.2 ± 8.5
Emotional & Behavioral Dyscontrol[a]	13.0 ± 4.8	14.2 ± 5.7	16.6 ± 6.0	18.9 ± 7.3	19.0 ± 6.8	16.1 ± 6.4
Sleep Disturbance[a]	13.1 ± 3.7	15.4 ± 5.1	18.2 ± 5.5	22.6 ± 6.2	21.9 ± 8.8	17.9 ± 6.2
Communication[b]	23.9 ± 2.1	23.3 ± 2.7	21.9 ± 3.5	19.3 ± 4.4	17.0 ± 5.4	21.9 ± 3.8
Executive Function[b]	38.0 ± 4.6	37.6 ± 3.8	34.7 ± 6.1	29.6 ± 7.9	24.5 ± 11.0	34.7 ± 6.7
General Cognitive Concerns[b]	34.4 ± 7.5	32.4 ± 7.1	28.1 ± 8.5	22.2 ± 9.3	20.5 ± 10.4	28.6 ± 9.2
Positive Affect & Well-Being[b]	40.0 ± 7.0	37.4 ± 5.8	34.3 ± 6.4	30.2 ± 7.0	25.8 ± 8.3	34.8 ± 7.1
Social Roles & Activities[b]	38.0 ± 4.1	35.7 ± 5.6	30.9 ± 6.8	25.2 ± 5.8	19.7 ± 6.2	31.7 ± 7.5
Lower Extremity Function[b]	38.4 ± 4.4	37.9 ± 4.8	34.2 ± 7.1	29.9 ± 6.7	19.7 ± 8.9	34.6 ± 7.1
Upper Extremity Function[b]	39.4 ± 1.9	39.4 ± 1.9	37.9 ± 3.9	35.5 ± 5.0	26.6 ± 10.3	37.8 ± 4.2

Note: All relationships with short forms were significant at $p < 0.001$.
[a]Higher score indicates worse functioning.
[b]Higher score indicates better functioning.

Table 6. Known-group analysis of variance for Neuro-QoL short forms raw scores with bowel and bladder scales.

	Bowel scale[a]			Bladder scale[a]		
Neuro-QoL short forms	Low $n = 550$	High $n = 225$	Total $n = 775$	Low $n = 480$	High $n = 301$	Total $n = 781$
Anxiety[a]	15.4 ± 6.8	19.6 ± 7.3	16.6 ± 7.1	14.9 ± 6.5	19.5 ± 7.3	16.7 ± 7.2
Depression[a]	12.8 ± 5.6	16.2 ± 6.7	13.8 ± 6.2	12.5 ± 5.4	16.0 ± 6.9	13.8 ± 6.2
Fatigue[a]	20.0 ± 8.0	26.4 ± 7.6	21.9 ± 8.4	19.8 ± 7.9	25.7 ± 8.0	22.1 ± 8.4
Emotional & Behavioral Dyscontrol[a]	14.8 ± 5.8	18.1 ± 6.5	15.8 ± 6.2	14.8 ± 5.6	17.9 ± 6.8	16.0 ± 6.3
Sleep Disturbance[a]	16.5 ± 5.6	20.4 ± 6.2	17.6 ± 6.0	16.4 ± 5.4	19.9 ± 6.3	17.7 ± 6.0
Communication[b]	22.9 ± 3.0	20.3 ± 4.4	22.1 ± 3.7	23.1 ± 2.9	20.4 ± 4.4	22.0 ± 3.8
Executive Function[b]	36.4 ± 5.2	31.5 ± 8.0	35.0 ± 6.5	36.6 ± 5.1	32.0 ± 7.8	34.8 ± 6.7
General Cognitive Concerns[b]	30.8 ± 8.4	24.4 ± 9.2	29.0 ± 9.1	31.1 ± 8.3	25.2 ± 9.5	28.8 ± 9.2
Positive Affect & Well-Being[b]	36.1 ± 6.6	32.2 ± 6.9	35.0 ± 6.9	36.4 ± 6.3	32.7 ± 7.2	35.0 ± 6.9
Social Roles & Activities[b]	34.0 ± 6.5	27.6 ± 7.5	32.1 ± 7.3	34.6 ± 6.1	28.0 ± 7.5	32.0 ± 7.4
Lower Extremity Function[b]	36.7 ± 5.5	31.0 ± 8.1	35.0 ± 6.8	37.0 ± 5.3	31.6 ± 7.7	34.9 ± 6.9
Upper Extremity Function[b]	38.9 ± 2.7	36.0 ± 6.1	38.0 ± 4.2	38.8 ± 3.0	36.9 ± 5.2	38.0 ± 4.1

Note: All relationships with Short Forms were significant at $p < 0.001$.
[a]Higher score indicates worse functioning.
[b]Higher score indicates better functioning.

Table 7. Summary of fit indices.

Short forms scale	RMSEA [90% CI]	TLI	CFI
Anxiety	0.144 [0.132, 0.156]	0.913	0.938
Depression	0.122 [0.110, 0.134]	0.939	**0.957**
Communication	0.079 [0.054, 0.105]	**0.976**	**0.988**
Executive Function	0.173 [0.161, 0.185]	0.876	0.911
General Cognitive Concerns	0.120 [0.108, 0.133]	**0.958**	**0.970**
Fatigue	0.180 [0.168, 0.192]	0.900	0.928
Executive & Behavioral Dyscontrol	0.163 [0.151, 0.176]	0.891	0.922
Positive Affect & Well-Being	0.202 [0.191, 0.213]	0.833	0.875
Social	0.161 [0.149, 0.174]	0.920	0.943
Sleep Disturbance	0.128 [0.115, 0.140]	0.843	0.888
Lower Extremity Function	0.141 [0.129, 0.154]	0.926	0.947
Upper Extremity Function	0.188 [0.175, 0.200]	0.827	0.877

Note: RMSEA = root mean square error of approximation; TLI = Tucker–Lewis Index; CFI = Comparative Fit Index.
Using recommended cutoffs for these indices, good fit is indicated by RMSEA < 0.06, TLI ≥ 0.95, and CFI ≥ 0.95.[26] Values satisfying recommended cutoffs are emphasized with bold-faced type.

General Cognitive Concerns: 0.970). The Depression SF had an acceptable CFI value (0.957) but no other indices were within the acceptable range.

Discussion

The current study sought to examine the discriminant validity of 12 Neuro-QoL SFs in a clinical population with MS. More specifically, we examined the reliability, validity, and factor structure of these measures. Overall, the Neuro-QoL SFs had acceptable reliability and validity. Consistent with findings reported by Miller and colleagues,[12] all 12 domains demonstrated significant and acceptable internal consistency. Neuro-QoL correlations with a measure of MS disease severity (PDDS) ranged from the nominal to the large range. Notably,

strong convergent validity was only evidenced in the Lower Extremity Function SF. This was expected given that the PDDS is largely focused on gait. All 12 SF scales were associated with a general quality-of-life measure (Question 1 from the SF-36). Similarly, known-group analysis with measures of bowel and bladder functioning showed worse Neuro-QoL outcomes were associated with worse bowel and bladder functioning. The direction of relationships between the SFs and the other measures (i.e., PDDS, SF-36, bowel and bladder function) were as would be expected such that greater disease severity and symptoms were associated with worse quality of life.

Questions remain with regards to the original factor structure of the Neuro-QoL in this clinical sample of MS patients. Results from the CFA suggest inadequate fit of the theoretical framework to the data reported here. The Communication and the General Cognitive Concerns scales were most strongly supported in our sample, while all other SFs demonstrated poor model fit. Notably, disagreement exists on the thresholds used for assessing model fit in CFA, which can affect comparability across studies. Specifically, issues related to other characteristics of the data beyond dimensionality have been shown to have an impact on CFI[32,33] and RMSEA.[33] In a recent analysis of similar patient-report outcome data, researchers used less restrictive thresholds to assess adequate model fit, namely RMSEA ≤ 0.08, TLI ≥ 0.95, and CFI ≥ 0.90.[34] However, application of these cutoffs to the present data would not have altered our findings supporting good fit only in the Communication and General Cognitive Concerns scales.

In light of research examining model fit indices, we agree that these cutoffs are not canonical benchmarks and should be approached as informative guides for additional exploration.[34] In addition to numerous factors, one explanation for why observed data can fail to fit a single factor model may be related to failure of the observed data to conform to the strict linear model assumptions used in factor analysis. The non-linearity in the observed data could be due to various factors, including the possibility that item responses are conditional upon each other, whereby respondents who answer in a particular range on one item may not endorse another item. For example, on the Neuro-QoL Anxiety SF, respondents who endorse "Never" for "Many situations made me worry" would likely respond with "Never" on "My worries overwhelmed me." Similarly, on the Fatigue SF, responses to several items are contingent on the response to the item "I felt tired." Some available measures and structured interviews in neurology account for this limitation by using a simple decision tree by which follow-up questions are only asked if an anchor item is endorsed. The Neuro-QoL Communication SF, which had good fit on all three indices, is also the most parsimonious of the 12 SFs evaluated with only five items compared with the eight or nine items on the other SFs. Moreover, qualitative review of the Communication SF items suggests the items are not conditional on each other; that is, while all items are related to an underlying construct, the response to any single item is not contingent on the response of another. These factors may help explain the strong unidimensional model fit observed in the Communication scale.

Overall, these findings suggest that, in MS, Neuro-QoL may have significant utility in evaluating and tracking PROs related to subjective communication and cognitive concerns. However, in this specific clinical population with MS, this system of PROs may have limited interpretability when evaluating other domains. In light of our CFA results, further work is needed to validate any justifiable modifications to improve the SFs.

This study is not without limitations. In contrast to previous research examining convergent validity of Neuro-QoL in MS,[12] we used fewer and shorter measures related to MS symptoms. This may have had an impact on the magnitude of the significant relationships observed. Despite this, Spearman's rho and ANOVA values were within acceptable ranges and consistent with previous work in this area.[12]

For the purpose of collecting PROs in MS, further exploration for an abridged version of the Neuro-QoL SFs may be warranted. It is possible, as suggested by the current findings, that fewer items or domains may be used to validly capture quality-of-life data from MS patients with the Neuro-QoL SFs.

Acknowledgements

The authors would like to thank the patients, staff, and clinicians at the Rocky Mountain Multiple Sclerosis Center as well as Richard N. Jones for his assistance with the interpretation of the results.

Conflict of Interests

The author(s) declared the following potential conflicts of interest with respect to the research, authorship, and/or

publication of this article: Enrique Alvarez has received research funding from Genentech, Biogen, Novartis, and the Rocky Mountain MS Center, as well as consulting fees from Genzyme, Genentech, Novartis, Acorda, Actelion, and Biogen. Kavita V. Nair has received research funding from Novartis, Biogen, Gilead Sciences as well as consulting fees from Astellas and Genentech. Luis D. Medina, Stephanie Torres, and Brooke Valdez have no relevant disclosures.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

ORCID iD

Luis D Medina https://orcid.org/0000-0001-8163-7772

Supplemental Material

Supplemental material for this article is available online.

References

1. Schaeffer J, Cossetti C, Mallucci G, et al. Multiple sclerosis. In: *Neurobiology of brain disorders*. Amsterdam: Elsevier, 2015, pp. 497–520.
2. Thompson AJ, Baranzini SE, Geurts J, et al. Multiple sclerosis. *Lancet* 2018; 391: 1622–1636.
3. Fowler. Lessons learned while integrating patient-reported outcomes in a psychiatric hospital. *Psychotherapy* 2018; 56: 91–99.
4. Talib TL, DeChant P, Kean J, et al. A qualitative study of patients' perceptions of the utility of patient-reported outcome measures of symptoms in primary care clinics. *Qual Life Res* 2018; 27: 3157–3166.
5. Baker A. Crossing the quality chasm: a new health system for the 21st century. *BMJ* 2001; 323: 1192.
6. Brundage M, Bass B, Jolie R, et al. A knowledge translation challenge: clinical use of quality of life data from cancer clinical trials. *Qual Life Res* 2011; 20: 979–985.
7. El Gaafary M. A guide to PROMs methodology and selection criteria. In: El Miedany Y (ed.) *Patient reported outcome measures in rheumatic diseases*. Cham: Springer International Publishing, 2016, pp. 21–58.
8. Ishaque S, Karnon J, Chen G, et al. A systematic review of randomised controlled trials evaluating the use of patient-reported outcome measures (PROMs). *Qual Life Res* 2019; 28: 567–592.
9. Victorson D, Cavazos JE, Holmes GL, et al. Validity of the Neurology Quality-of-Life (Neuro-QoL) measurement system in adult epilepsy. *Epilepsy Behav* 2014; 31: 77–84.
10. Cella D, Gershon R, Lai J-S, et al. The future of outcomes measurement: item banking, tailored short-forms, and computerized adaptive assessment. *Qual Life Res* 2007; 16: 133–141.
11. Cella D, Lai J-S, Nowinski CJ, et al. Neuro-QOL: brief measures of health-related quality of life for clinical research in neurology. *Neurology* 2012; 78: 1860–1867.
12. Miller DM, Bethoux F, Victorson D, et al. Validating Neuro-QoL short forms and targeted scales with people who have multiple sclerosis. *Mult Scler* 2016; 22: 830–841.
13. Farrell AM and Rudd J. Factor analysis and discriminant validity: a brief review of some practical issues. In: Proceedings of the Australian and New Zealand Marketing Academy (ANZMAC) Sustainable Management and Marketing Conference (ed D Tojib), Melbourne, Australia, 30 November–02 December 2009, paper no. 408960, pp. 1–7. Australia and New Zealand: ANZMAC.
14. Zait A and Bertea PE. Methods for testing discriminant validity. *Manag Mark J* 2011; IX: 217–224.
15. Brown JD. Questions and answers about language testing statistics: how are PCA and EFA used in language research? *JALT* 2010; 14: 19–23.
16. McDonald WI, Compston A, Edan G, et al. Recommended diagnostic criteria for multiple sclerosis: guidelines from the international panel on the diagnosis of multiple sclerosis. *Ann Neurol* 2001; 50: 121–127.
17. Hohol MJ, Orav EJ and Weiner HL. Disease Steps in multiple sclerosis: a simple approach to evaluate disease progression. *Neurology* 1995; 45: 251–255.
18. Learmonth YC, Motl RW, Sandroff BM, et al. Validation of patient determined disease steps (PDDS) scale scores in persons with multiple sclerosis. *BMC Neurol* 2013; 13. DOI: 10.1186/1471-2377-13-37.
19. Ware JE and Gandek B. Methods for testing data quality, scaling assumptions, and reliability: the IQOLA Project approach. *J Clin Epidemiol* 1998; 51: 945–952.
20. Muthén LK and Muthén BO. *Mplus user's guide*. 6th ed. Los Angeles: Muthén & Muthén, 2010.
21. IBM Corp. *SPSS Statistics, Version 25.0*. Armonk: IBM Corp., 2017.
22. Peterson RA. A meta-analysis of Cronbach's coefficient alpha. *J Consum Res* 1994; 21: 381–391.
23. Ware JE Jr and Sherbourne CD. The MOS 36-Item Short-Form Health Survey (SF-36): I. conceptual framework and item selection. *Med Care* 1992; 30: 473–483.
24. Hurley AE, Scandura TA, Schriesheim CA, et al. Exploratory and confirmatory factor analysis: guidelines, issues, and alternatives. *J Organ Behav* 1997; 18: 667–683.
25. Schreiber JB, Nora A, Stage FK, et al. Reporting structural equation modeling and confirmatory factor analysis results: a review. *J Educ Res* 2006; 99: 323–338.
26. Hu LT and Bentler PM. Evaluating model fit. In: Hoyle RH (ed.) *Structural equation modeling:*

Concepts, issues, and application. Thousand Oaks: Sage, pp. 77–99.
27. Yong AG and Pearce S. A beginner's guide to factor analysis: focusing on exploratory factor analysis. *Tutor Quant Methods Psychol* 2013; 9: 79–94.
28. Field AP, Miles J and Field Z. *Discovering statistics using R*. London: Sage, 2012.
29. Little RJA. A test of missing completely at random for multivariate data with missing values. *J Am Stat Assoc* 1988; 83: 1198–1202.
30. Fichman M and Cummings JN. Multiple imputation for missing data: making the most of What you know. *Organ Res Methods* 2003; 6: 282–308.
31. Menard SW. *Applied logistic regression analysis*. Thousand Oaks: Sage Publications, 1995.
32. Reise SP, Scheines R, Widaman KF, et al. Multidimensionality and structural coefficient bias in structural equation modeling: a bifactor perspective. *Educ Psychol Meas* 2013; 73: 5–26.
33. Cook KF, Kallen MA and Amtmann D. Having a fit: impact of number of items and distribution of data on traditional criteria for assessing IRT's unidimensionality assumption. *Qual Life Res* 2009; 18: 447–460.
34. Cook KF, Kallen MA, Bombardier C, et al. Do measures of depressive symptoms function differently in people with spinal cord injury versus primary care patients: the CES-D, PHQ-9, and PROMIS®-D. *Qual Life Res* 2017; 26: 139–148.