

# Neuro-QOL

## Brief measures of health-related quality of life for clinical research in neurology



D. Cella, PhD  
J.-S. Lai, PhD  
C.J. Nowinski, MD, PhD  
D. Victorson, PhD  
A. Peterman, PhD  
D. Miller, PhD  
F. Bethoux, MD  
A. Heinemann, PhD  
S. Rubin, MD  
J.E. Cavazos, MD, PhD  
A.T. Reder, MD  
R. Sufit, MD  
T. Simuni, MD  
G.L. Holmes, MD  
A. Siderowf, MD  
V. Wojna, MD  
R. Bode, PhD  
N. McKinney  
T. Podrabsky  
K. Wortman, MSW  
S. Choi, PhD  
R. Gershon, PhD  
N. Rothrock, PhD  
C. Moy, PhD

Correspondence & reprint requests to Dr. Cella: d-cella@northwestern.edu

Editorial, page 1813

Supplemental data at [www.neurology.org](http://www.neurology.org)

Supplemental Data



Podcast



### ABSTRACT

**Objective:** To address the need for brief, reliable, valid, and standardized quality of life (QOL) assessment applicable across neurologic conditions.

**Methods:** Drawing from larger calibrated item banks, we developed short measures (8–9 items each) of 13 different QOL domains across physical, mental, and social health and evaluated their validity and reliability. Three samples were utilized during short form development: general population (Internet-based,  $n = 2,113$ ); clinical panel (Internet-based,  $n = 553$ ); and clinical outpatient (clinic-based,  $n = 581$ ). All short forms are expressed as T scores with a mean of 50 and SD of 10.

**Results:** Internal consistency (Cronbach  $\alpha$ ) of the 13 short forms ranged from 0.85 to 0.97. Correlations between short form and full-length item bank scores ranged from 0.88 to 0.99 (0.82–0.96 after removing common items from banks). Online respondents were asked whether they had any of 19 different chronic health conditions, and whether or not those reported conditions interfered with ability to function normally. All short forms, across physical, mental, and social health, were able to separate people who reported no health condition from those who reported 1–2 or 3 or more. In addition, scores on all 13 domains were worse for people who acknowledged being limited by the health conditions they reported, compared to those who reported conditions but were not limited by them.

**Conclusion:** These 13 brief measures of self-reported QOL are reliable and show preliminary evidence of concurrent validity inasmuch as they differentiate people based upon number of reported health conditions and whether those reported conditions impede normal function. *Neurology*<sup>®</sup> 2012;78:1860–1867

### GLOSSARY

**ALS** = amyotrophic lateral sclerosis; **CAT** = computerized adaptive test; **IRT** = item response theory; **MS** = multiple sclerosis; **NINDS** = National Institute of Neurological Disorders and Stroke; **PD** = Parkinson disease; **QOL** = quality of life.

In neurology clinical research, traditional outcome measures of disease status often fail to represent the full impact of disease and treatment. The patient's experience of disease symptoms, treatment side effects, functioning, and well-being—commonly referred to as health-related quality of life (QOL)—is often not included in a systematic evaluation of clinical benefit. Yet, the patient's experience of disease and treatment can be the key driver of treatment impact, acceptability, or value.<sup>1</sup> While many QOL scales are available to the neurology clinical researcher, some have questionable validity or may be difficult to interpret. In addition, different instruments tend to be used in different neurologic conditions, rendering cross-disease evaluations of QOL burden or benefit impossible.<sup>2–6</sup> Even within a given condition, there is seldom consensus on common measures, which impedes cross-study comparisons of relative disease burden, benefits of different treatments, or other factors.

From the Northwestern University Feinberg School of Medicine (D.C., J.-S.L., C.J.N., D.V., A.H., R.S., T.S., R.B., N.M., T.P., K.W., S.C., R.G., N.R.), Chicago, IL; University of North Carolina—Charlotte (A.P.), Charlotte; Cleveland Clinic Foundation (D.M., F.B.), Cleveland, OH; Rehabilitation Institute of Chicago (A.H.), Chicago, IL; Northshore University Health System (S.R.), Evanston, IL; University of Texas Health Science Center at San Antonio (J.E.C.), San Antonio; University of Chicago (A.T.R.), Chicago, IL; Dartmouth Medical School (G.L.H.), Hanover, NH; University of Pennsylvania (A.S.), Philadelphia; Medical Sciences Campus (V.W.), University of Puerto Rico, San Juan; and National Institute of Neurological Disorders and Stroke (C.M.), Bethesda, MD.

**Study funding:** The Neuro-QOL project was supported by the National Institute of Neurological Disorders and Stroke (NINDS) contract number HHSN265200423601C (PI: David Cella).

Go to [Neurology.org](http://Neurology.org) for full disclosures. Disclosures deemed relevant by the authors, if any, are provided at the end of this article.

In an effort to address these limitations, the National Institute of Neurologic Disorders and Stroke (NINDS) sponsored a multisite project to develop a clinically relevant and psychometrically robust QOL assessment tool for adults and children with neurologic disorders.<sup>7,8</sup> This effort, Neuro-QOL, enables clinical researchers to compare the QOL impact of different interventions within and across various conditions. In this article, we summarize the development and validation of the first generation of 13 brief adult Neuro-QOL short forms for use in clinical neurology research.

**METHODS Constructing item banks.** To build Neuro-QOL item banks, we followed a series of steps designed to ensure clinical and psychometric validity. These steps included identifying the needs of the clinical research community,<sup>8,9</sup> ensuring clinical and patient-driven evidence of importance and relevance of the selected QOL domains, and an expert consensus-based selection of priority conditions.<sup>10</sup> We combined input from patient and caregiver focus groups<sup>11,12</sup> with expert input and a literature review, to determine the QOL domains to include in Neuro-QOL.<sup>8</sup> We then conducted large-scale testing to calibrate item response theory (IRT)-based<sup>13</sup> item banks across physical, mental, and social domains of QOL.<sup>14,15</sup> Each Neuro-QOL bank includes a large collection of items (questions and their response options) that have been evaluated and tested to ensure their relevance, clarity, fit with the concept being measured, and informativeness.<sup>15</sup> This produced 13 adult QOL item banks (anxiety; depression; fatigue; upper extremity function—fine motor, activities of daily living; lower extremity function—mobility; applied cognition—general concerns; applied cognition—executive function; emotional and behavioral dyscontrol; positive affect and well-being; sleep disturbance; ability to participate in social roles and activities; satisfaction with social roles and activities; stigma).

Neuro-QOL item banks enable researchers to select or design static short form measures or to administer a dynamic computerized adaptive test (CAT).<sup>16,17</sup> The Neuro-QOL CAT can be tailored to each respondent, selecting the most informative next question based on previous responses. In general, CATs provide the most precise estimate of patients' health status with the fewest number of questions as only the most informative items are selected iteratively.<sup>18</sup> A second option is the use of short forms, or subsets of questions from the bank. Short forms can be of any length, ranging from 1 question to 20 or more. In each case, the score generated for the respondent is expressed on a common metric or scale. For Neuro-QOL, we report scores using a T distribution, with the mean of the reference population set to 50 and the SD set to 10 units. Based upon the samples required to obtain stable item statistics, Neuro-QOL T scores are anchored either to the general US population (designated as "GPT") or to a clinical population (designated as "CT" scores).

**Standard protocol approvals, registrations, and patient consents.** All research activities reported here received Institutional Review Board approval and all participants provided informed consent.

**Samples and assessment procedure.** The samples studied for this work are described in detail elsewhere.<sup>14,15</sup> Briefly, we engaged 3 adult samples for item pool testing. The US general population sample included a total of 3,123 English-speaking and Spanish-speaking respondents recruited through an online panel company; data from the subset of 2,113 English speakers were used for calibrating item parameters and setting the central (average) location for each T score (mean = 50; SD = 10). This subset was 50% male (mean age = 52.7 years; SD = 15.5 years). These 2,113 participants were divided into 4 blocks of at least 500, with each block given 1 of 4 item pools: physical function, emotional health, social function, and cognitive function. A second clinical panel sample of 553 people with a physician-confirmed diagnosis of epilepsy, stroke, amyotrophic lateral sclerosis (ALS), multiple sclerosis (MS), or Parkinson disease (PD) was also recruited, by a different online panel company, to calibrate the stigma bank and disease-targeted scales (because such scales cannot be answered by people without a medical diagnosis). The clinical panel sample was 53% male with mean age of 56.2 years (SD = 12.8).<sup>15</sup> Finally, a third sample of 581 outpatient neurology patients was drawn from collaborating neurologists around the United States and Puerto Rico, and included patients with physician-diagnosed epilepsy, stroke, ALS, MS, or PD. Comorbidity was not an exclusionary criterion. This sample, the clinical outpatient sample, was 46% male (mean age = 55.21; SD = 14.3). They were included in the initial item calibration study<sup>11</sup> to obtain reliable calibrations for the physical function, applied cognition, and sleep item banks. Demographic details on each sample are provided in table 1.

We asked respondents in both online samples the following "yes/no" question: "Have you ever been told by a doctor or health professional that you have <condition/disease>?" We queried for 19 conditions: hypertension; chest pain; coronary artery disease; heart failure or congestive heart failure; heart attack; stroke; migraines or severe headaches; diabetes, high blood sugar, or sugar in urine; cancer (other than nonmelanoma skin cancer); depression; anxiety; alcohol or drug problem; sleep disorder; HIV or AIDS; spinal cord injury; MS; PD; epilepsy; and ALS. For each condition endorsed, we then asked: "Are any of your current activities limited by your <condition/disease>? (Yes or No)." Respondents were then sorted into 3 groups by number of reported conditions (0; 1–2 conditions; 3 or more conditions), and all respondents who reported at least 1 condition were sorted again within the above categories into "yes" or "no" activity limitation.

We hypothesized that respondents who reported more conditions, and those who reported activity limitation, would have worse QOL.

**Constructing and evaluating the short forms.** Each short form was constructed using the same approach. Starting with item statistics generated from the IRT item calibrations (response category threshold and slope parameters),<sup>15</sup> we ranked items by the amount of information they provided across the range of what was being measured (e.g., applied cognition). We also ran CAT simulations to identify items selected early in the procedure. Because the CAT algorithm weights information heavily in item selection, there was overlap between information ranks and CAT ranks, although some items were ranked highly in one but not the other criterion. Ten doctoral level clinical and measurement experts (3 neurologists; 4 clinical psychologists; 1 occupational therapist; 1 social worker; 1 neuropsychologist) then reviewed each candidate item for relevance and appeal based on item content only. Item performance (information and

**Table 1** Sample demographics

	Clinical panel sample	General population (English-speaking)	Clinical outpatient sample
No.	553	2,113	581
Age average (SD)	56.2 (12.8)	52.67 (15.5)	55.21 (14.3)
<b>Gender, %</b>			
Male	53	50	46
Female	47	50	54
<b>Race, %</b>			
White	95	91	87
African American	3	5.5	12
American Indian/Alaskan Native	4	1.5	2
Asian	1	3.3	2
Hawaiian/Pacific Islander	6	1.0	0
<b>Occupation, %</b>			
Homemaker	11.5	12	8
Unemployed	8	8	9
Retired	37	31	30
Disability/leave	32	11	35
Full-time employed	25	31	21
Part-time employed	10	12	10
Full-time student	2	3	1
<b>Marital status, %</b>			
Married	60	52	62
Separated/divorced	16	17	13
Widowed	7	7	5
Living with someone	6.5	7	5
Never married	11	17	16
<b>Income, %</b>			
>\$20,000	17	18	16
\$20-\$49,000	35	45	35
\$50-\$99,000	30.5	31	28
<\$100,000	14.5	11	21
<b>Education, %</b>			
Some high school or less	3.5	2	3
High school or equivalent	14.5	22	19
Some college	40	40	29
College degree	21	24	29
Advanced degree	22	11	20

CAT rank order) was not shared with experts prior to their ratings. Experts identified their 5 most-preferred and 5 least-preferred items in the calibrated bank. Individual preferences of each rater were then presented along with item performance statistics.

With the above information tabulated, we (D.C., D.V., S.C., J.-S.L., C.N., D.M., N.R.) identified items with strong psychometric characteristic (IRT model fit; highly informative; selected early by CAT) and high appeal to clinical raters. We discussed marginal item choices (e.g., high clinical appeal but relatively weak psychometric performance) until we reached consensus regarding item inclusion in the short form. We also con-

sidered 2 other goals: 1 was respondent burden, so that if 1 of 2 nearly equal items had the same response options as the other selected items, it was selected. The other was inclusion of items from the Patient-Reported Outcomes Measurement Information System (PROMIS; www.nihpromis.org), if they were calibrated with other items in the Neuro-QOL bank and were not ranked very low by either of the sources of input. This extra step was taken to maximize the probability that Neuro-QOL can be linked (cross-walked) to the PROMIS item banks.

Using the samples described above, we produced T scores for each of the 13 short forms (note that sleep disturbance is an 8-item bank so no short version was developed). We computed Cronbach  $\alpha$  coefficient to evaluate internal consistency, and correlated each short form score with the score derived from the full item bank (correcting for item overlap by removing short form items from the full bank prior to correlating scores). We evaluated the ability of each short form to 1) differentiate people who reported no health condition from those who reported 1–2 conditions or 3 or more conditions, and 2) differentiate people who reported activity limitation due to a health condition from those who reported no limitation. Finally, we compared the results of these 2 comparisons to results of the same comparisons using the corresponding full item banks.

**RESULTS** We began with 13 item banks (table 2). Applying the approach described above, we developed 8- or 9-item short forms for banks that included more than 10 items. We then calculated internal consistency of each short form using Cronbach  $\alpha$  coefficient, and evaluated correlations between each short form and its respective bank, using Spearman rank order correlation for ordinal data. However, because this association can be inflated by redundancy of items in both scales, especially with smaller banks, we also report Spearman correlations between short forms and banks after excluding items that comprised the actual short forms. These results can be found in table 2.

After establishing their internal consistency reliability and strong association with full bank scores, we examined the association of each short form with all other short forms. Our purpose was to ensure relatively higher correlations with related concepts (e.g., depression and anxiety, lower extremity and upper extremity physical function), and relatively lower correlations with unrelated concepts (e.g., depression and lower extremity function). These results are available in table 3.

To evaluate the validity of the 13 developed short forms, we tested their ability to differentiate subgroups of participants hypothesized to have poorer QOL. Using 13 separate one-way analyses of variance, we compared participants who reported 0, vs 1–2, vs 3 or more comorbid conditions. In 3 of the 13 comparisons (fatigue, emotional and behavioral dyscontrol, stigma) there were no respondents with “0” diagnoses reported, because this sample (the “clinical panel sample”) was selected on the basis of

**Table 2** Number of items and associations between Neuro-QOL adult item banks and short forms

Item bank	No. items in bank	No. items in short form	Sample n	Cronbach $\alpha$ for short form	Correlation between SF and bank (xx/yy <sup>a</sup> )
Anxiety	21	8	513	0.95	0.96/0.92
Depression	24	8	513	0.96	0.96/0.94
Fatigue	19	8	511	0.97	0.98/0.95
Upper extremity function: fine motor, ADL	20	8	1,094	0.93	0.88/0.83
Lower extremity function: mobility	19	8	1,043	0.93	0.95/0.89
Applied cognition: executive function	13	8	1,109	0.92	0.95/0.82
Applied cognition: general concerns	18	8	1,109	0.94	0.97/0.91
Emotional and behavioral dyscontrol	18	8	511	0.93	0.97/0.90
Positive affect and well-being	23	9	513	0.97	0.99/0.96
Sleep disturbance	8	8	1,087	0.85	NA <sup>b</sup>
Ability to participate in social roles and activities	45	8	549	0.96	0.95/0.94
Satisfaction with social roles and activities	45	8	549	0.96	0.96/0.94
Stigma	24	8	511	0.93	0.96/0.93

Abbreviations: ADL = activities of daily living; NA = not applicable; QOL = quality of life; SF = short form.

<sup>a</sup> xx = before removal of SF items from bank; yy = after removal of SF items from bank.

<sup>b</sup> Sleep disturbance is a small calibrated item bank of only 8 items so we did not develop a short form.

having a neurologic disorder (PD, MS, epilepsy, stroke, or ALS). In every analysis, a significant *F* value was obtained with mean differences between groups in the hypothesized direction (more conditions associated with worse QOL). Usually, the difference in mean score was greater between the 1–2 vs the 3 or more group than between the 0 and the 1–2 group (table 4).

Within each comorbidity group (1–2 and 3 or more), we next divided participants into those who reported activity limitation from their health condition vs those who did not. If a patient with multiple conditions reported activity limitation on only a subset of those reported conditions, he or she was counted as “yes” with regard to activity limitation from disease. The results of these comparisons, including *t* tests and effect sizes for group differences, are also reported in table 4. For each of the 2 comorbidity groups (1–2 and 3 or more), across all 13 short forms (i.e., 26 comparisons), the *t* test was significant. Effect sizes for the difference in score between those with vs without activity limitation from their health condition was moderate to very large (effect size range = 0.43–1.58). As a final check on the performance of the short forms relative to the full item banks, we plotted short form scores against full bank scores for every subgroup represented in table 4, and for the overall group. Correlations between short form and full bank scores ranged from 0.97 to 1.00 for each subgroup, and was 0.91 overall. This is plotted in the figure.

**DISCUSSION** We report on the development and initial validation of 13 brief measures of QOL for

adults with neurologic disorders. Each Neuro-QOL short form comprises a set of items that have been carefully selected from item banks to enhance estimation of a patient’s health status. The length of each short form ranges from 8 to 9 items; each can be completed in less than 2 minutes by the typical patient.<sup>19</sup> A profile of 6 selected domains, for example, would require approximately 10 minutes to complete. This compares favorably to other QOL instruments in common use, which can take as many as 20 minutes or more. Scoring look-up tables are available in appendix e-1 on the *Neurology*<sup>®</sup> Web site at [www.neurology.org](http://www.neurology.org). Researchers also may design their own short forms by selecting items from the item banks. In that case, scoring and converting to the T-score metric can be done with direction also provided in appendix e-1. The short forms reported here provide a practical opportunity for multidimensional assessment in neurologic clinical research or practice. Over time and with accumulating publications, their use can be enhanced by increased interpretability with regard to the meaning of specific scores and score changes. For now, the interpretability of these scales surrounds the reference point provided by the T score. Specifically, the following 4 banks were referenced against a clinical neurology population: sleep disturbance, fatigue, emotional and behavioral dyscontrol, and stigma. T scores from the other 9 item banks are referenced against the US general population. Therefore, when interpreting sleep disturbance, fatigue, emotional and behavioral dyscontrol, or stigma scores, one should consider a score

**Table 3** Correlations among Neuro-QOL short forms<sup>a</sup>

	Depression	Anxiety	Stigma	Positive affect and well-being	Applied cognition: General concerns	Applied cognition: Executive function	Lower extremity function: Mobility	Upper extremity function: Fine motor, activities of daily living	Ability to participate in social roles and activities	Satisfaction with social roles and activities	Fatigue	Sleep disturbance
Depression	1.000											
Anxiety	0.711	1.000										
Stigma	0.589	0.514	1.000									
Positive affect and well-being	-0.723	-0.565	-0.485	1.000								
Applied cognition: General concerns	-0.356	-0.424	-0.311	0.349	1.000							
Applied cognition: Executive function	-0.310	-0.320	-0.284	0.324	0.692	1.000						
Lower extremity function: Mobility	-0.240	-0.226	-0.383	0.295	0.200	0.245	1.000					
Upper extremity function: Fine motor, activities of daily living	-0.234	-0.221	-0.404	0.256	0.144	0.267	0.666	1.000				
Ability to participate in social roles and activities	-0.485	-0.451	-0.527	0.594	0.329	0.366	0.429	0.438	1.000			
Satisfaction with social roles and activities	-0.527	-0.468	-0.560	0.572	0.308	0.269	0.449	0.447	0.722	1.000		
Fatigue	0.563	0.567	0.473	-0.540	-0.416	-0.351	-0.350	-0.280	-0.514	-0.508	1.000	
Sleep disturbance	0.517	0.569	0.483	-0.452	-0.446	-0.346	-0.263	-0.227	-0.381	-0.382	0.653	1.000
Emotional and behavioral dyscontrol	0.561	0.615	0.527	-0.439	-0.454	-0.348	-0.265	-0.255	-0.392	-0.411	0.507	0.568

Abbreviation: QOL = quality of life.

<sup>a</sup> All correlations were statistically significant at  $p < 0.001$ . Sample sizes of each comparison ranged from 597 (upper extremity function: fine motor, activities of daily living vs emotional and behavioral dyscontrol) to 663 (upper extremity function: fine motor, activities of daily living vs applied cognition: executive function).

**Table 4** Neuro-QOL short form scores on calibration sample and subgroup comparisons by number of health conditions and activity limitation (4 of 13 short forms shown; see table e-1 for complete table)<sup>a</sup>

Short form	No.	Mean (SD)	No. of reported diagnoses		F statistic <sup>b</sup>	Activity limitation within each group			
			Group (n)	Mean (SD)		Limitation (n)	Mean (SD)	t Test <sup>c</sup>	Effect size
Lower extremity function: Mobility	518	48.5 (7.9)	0 (189)	52.7 (5.6)	$F_{2,515} = 56.9$				
			1-2 (190)	47.5 (7.8)		Yes (48)	44.5 (8.0)	3.15 <sup>d</sup>	-0.53
						No (142)	48.5 (7.5)		
			3+ (139)	44.3 (8.1)		Yes (106)	43.2 (8.2)	2.86 <sup>d</sup>	-0.57
					No (33)	47.8 (7.0)			
Applied cognition: Executive function	533	49.9 (8.4)	0 (144)	53.0 (6.3)	$F_{2,530} = 36.0$				
			1-2 (218)	51.2 (7.5)		Yes (46)	48.2 (7.8)	3.12 <sup>d</sup>	-0.52
						No (172)	52.0 (7.2)		
			3+ (171)	45.8 (9.5)		Yes (117)	44.0 (9.4)	3.92 <sup>d</sup>	-0.65
					No (54)	49.9 (8.4)			
Ability to participate in social roles and activities	549	48.6 (7.4)	0 (189)	51.7 (6.1)	$F_{2,546} = 68.8$				
			1-2 (224)	49.2 (6.6)		Yes (56)	44.1 (6.2)	7.44 <sup>d</sup>	-1.15
						No (168)	50.9 (5.8)		
			3+ (136)	43.2 (7.3)		Yes (102)	40.9 (6.1)	7.39 <sup>d</sup>	-1.47
					No (34)	49.9 (6.3)			
Depression	513	47.8 (8.7)	0 (138)	44.2 (7.2)	$F_{2,510} = 38.4$				
			1-2 (193)	46.6 (8.0)		Yes (62)	50.5 (8.5)	-4.91 <sup>d</sup>	0.76
						No (131)	44.8 (7.1)		
			3+ (182)	51.9 (8.8)		Yes (122)	54.4 (7.8)	-6.18 <sup>d</sup>	0.98
					No (60)	46.7 (8.3)			

Abbreviation: QOL = quality of life.

<sup>a</sup> Neuro-QOL uses a T score which has a mean of 50 and SD of 10, based on the norming sample used. All Neuro-QOL banks and scales are scored such that a high score reflects more of what is being measured. Therefore, a high score is good for the first 3 of the 4 domains above, whereas a high score is bad for the last domain above (depression). A complete table 4 with all 13 short forms can be found as table e-1 on the *Neurology*<sup>®</sup> Web site at [www.neurology.org](http://www.neurology.org) and at [www.neuroqol.org](http://www.neuroqol.org).

<sup>b</sup> Analysis of variance between number of diagnoses reported by samples;  $p < 0.001$  for all comparisons.

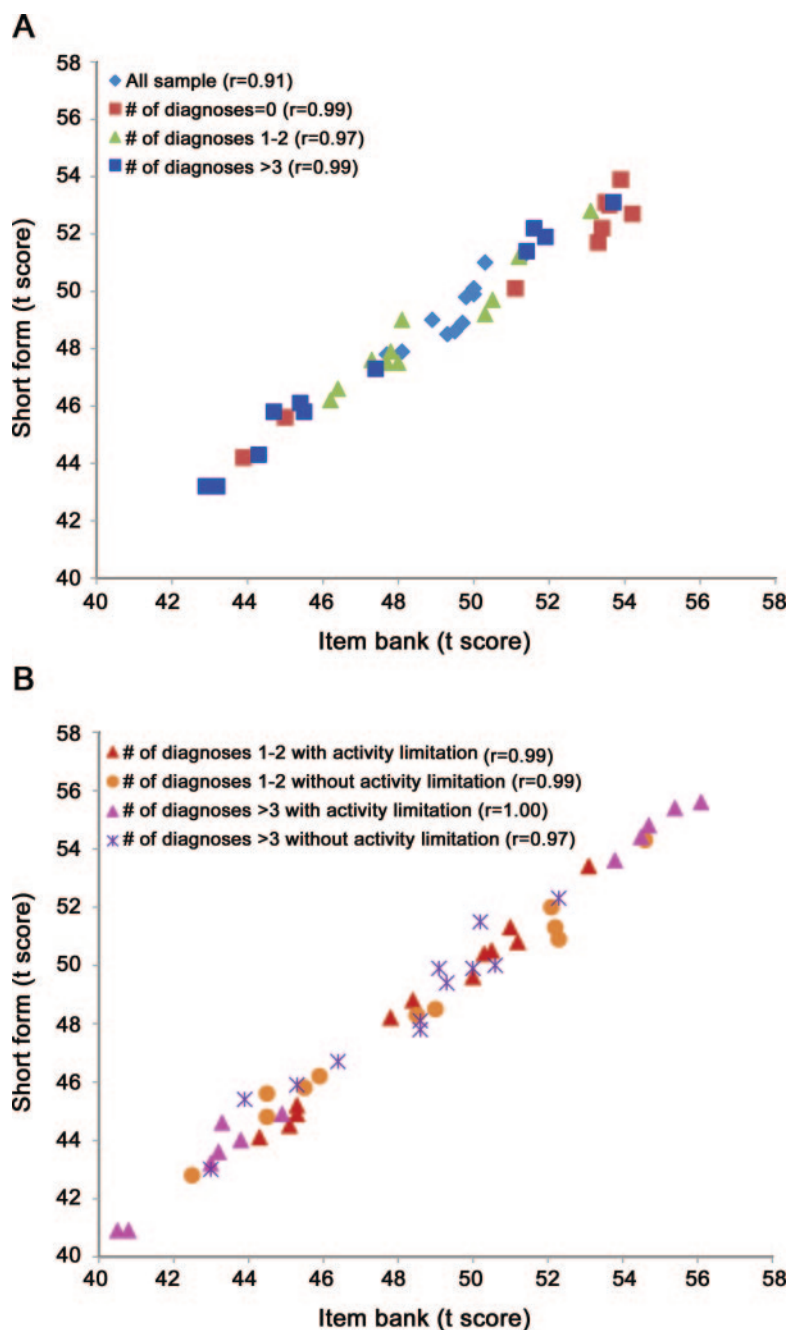
<sup>c</sup> t Test between samples who reported yes vs no activity limitation: <sup>d</sup>  $p < 0.01$ .

of 60, for example, to be 1 SD higher (worse) than the average of the clinical neurology sample described here and elsewhere.<sup>11</sup> When interpreting scores on the other 9 item banks or short forms, one should consider the reference group to be the US general population. That same score of 60 would be 1 SD higher (worse) than the average US resident (rather than neurology patient) on anxiety or depression, and 1 SD higher (better) than the average US resident on upper and lower extremity function, applied cognition, positive affect and well-being, ability to participate in social roles and activities, and satisfaction with social roles and activities.

The Neuro-QOL measurement system is intended to be brief, reliable, valid, responsive, and consistent enough across the selected conditions to allow for cross-disease comparison, and yet flexible enough to capture condition-specific HRQOL issues. However, there are limitations in the current

work which can be addressed in future research. First, the calibration samples were essentially samples of convenience, with most respondents recruited through Internet panel companies. The impact of this sampling strategy is likely negligible with regard to the integrity of the item statistics (“calibrations”), because what is most important for calibration is obtaining a full range of responses to items administered. When the general population sample did not provide sufficient responses in the most impaired response option (i.e., physical function, applied cognition, and sleep), we supplemented cases from the clinical outpatient sample to obtain stable item parameter statistics. However, the predominant use of an Internet panel sample raises questions about the generalizability of the results and the interpretation of T scores. Further research with populations that are not regular Internet users and those with limited reading ability will be important. In addition, it will

**Figure** Correlations between short form and full item bank scores within subgroups and overall by (A) number of diagnosis and (B) activity limitation



Correlations are given in parentheses after each group listed in the keys.

be very important to evaluate use of these short forms with patients who have limited functional or expressive ability and with proxies. Finally, although we developed Spanish language equivalent assessments for all of these 13 QOL domains, they have not been formally tested or evaluated.

Standardized QOL evaluations such as Neuro-QOL can inform health care accountability, from patient care to health care policy. It does so by improving assessment of patient-reported outcomes

and disease burden in neurologic diseases, increasing measurement consistency across neurologic clinical research, and offering a common metric to express burdens of disease and benefits of treatment. Over time, accumulated experience and published results with Neuro-QOL measures will support their use in a variety of applications, from clinical trial research to broader comparative effectiveness research, cross-sectional and longitudinal observational cohort studies, health care delivery observational and intervention studies, and population-based research.

### AUTHOR CONTRIBUTIONS

Statistical analyses were conducted by R. Bode, S. Choi, J.-S. Lai, N. McKinney, and T. Podrabsky.

### ACKNOWLEDGMENT

The authors thank their project manager, Vitali Ustsinovich, MA.

### DISCLOSURE

Dr. Cella has received research support from the National Institute of Neurological Disorders and Stroke (NINDS) contract number HHSN265200423601C. Dr. Lai reports no disclosures. Dr. Nowinski receives or has received research support from the NIH (contracts #HHSN265200423601C and #HHSN260200600007C) and Teva Pharmaceuticals. She has also received honoraria for writing an article for Medlink. Dr. Victorson holds stock options in Eli Lilly and Company, received an honoraria for serving on the Steering Committee of the Reeve Neuro-Recovery Network, was funded by NIH contracts #HHSN265200423601C and #HHS-N-260-2006-00007-C and grants #R01HD054569-02NIDRR, #1U01NS056975-01, and #R01CA104883, received support from the American Cancer Society (national and Illinois Division) for research in prostate cancer, received institutional support from NorthShore University HealthCare System for research in prostate cancer, received institutional support from the Medical University of South Carolina for sarcoidosis research, and received institutional support from the Northwestern Medical Faculty Foundation for urology research. Dr. Peterman receives royalties from an online entry in UpToDate; was supported by NIH Contract #HHSN265200423601C for the study on which this manuscript reports; and received institutional funding for research on spirituality and health. Dr. Miller has received research support from NIH contract HHSN26520043601C and TEVA Neuro-Sciences. She has received consulting fees from Biogen Idec and the Consortium of Multiple Sclerosis Centers. She serves on the editorial board of *Quality of Life Research*. Dr. Bethoux receives honoraria for consulting from Medtronic Inc., Biogen Idec, and IMPAX Laboratories. Dr. Heineemann receives salary support from a variety of federal research and training grants: Enhancing Quality of Prosthetic and Orthotic Services with Process and Outcome Information, National Institute on Disability and Rehabilitation Research (H133E080009); Rehabilitation Research and Training Center on Improving Measurement of Medical Rehabilitation Outcomes (H133B090024), National Institute on Disability and Rehabilitation Research; Midwest Regional Spinal Cord Injury Care System (H133N060014), National Institute on Disability and Rehabilitation Research; Midwest Regional Traumatic Brain Injury Care System, National Institute on Disability and Rehabilitation Research (H133A080045); Development of Quality Measures for Post-Stroke Rehabilitation, NIDRR/USDE. Dr. Rubin, Dr. Cavazos, and Dr. Reder report no disclosures. Dr. Sufit has served on Data Safety and Monitoring Boards for the NINDS and Pfizer Pharmaceuticals. He has received honoraria for speaking engagements from Hill-Rom. He has served as an expert in medical malpractice litigation. Dr. Simuni reports no disclosures. Dr. Holmes serves on the Advisory Board for Questcor Pharmaceuticals, Sunovion Pharmaceuticals, and Upsher-Smith Laboratories. Dr. Siderowf is supported by a Morris K. Udall Parkinson's Disease Research Center of Excellence grant from NINDS (NS-053488), and has been supported by SAP4100027296, a health research grant awarded by the Department of

Health of the Commonwealth of Pennsylvania from the Tobacco Master Settlement Agreement under Act 2001-77. Dr. Wojna has received personal compensation for activities with GlaxoSmithKline as a speaker and research support from Biogen Idec, and is funded by NIH grants #2U54NS43011, #U54RR022762 (pilot study), and #S11NS46278. Dr. Bode, N. McKinney, T. Podrabsky, K. Wortman, and Dr. Choi report no disclosures. Dr. Gershon has received personal compensation for activities as a speaker and consultant with Sylvan Learning, Rockman, and the American Board of Podiatric Surgery. He has several grants awarded by NIH: N01-AG-6-0007, 1U5AR057943-01, HHSN260200600007, 1U01DK082342-01, AG-260-06-01, HD05469, NINDS: U01 NS 056 975 02, NHLBI K23: K23HL085766, NIA: 1RC2AG036498-01, NIDRR: H133B090024, OppNet: N01-AG-6-0007 (PI: David Cella). Dr. Rothrock receives royalties from the publication of "Evaluation of Health-Related Quality of Life" in UpToDate. She receives research support from the National Institutes of Health and previously the Center for Disease Control and Prevention. Dr. Moy reports no disclosures. **Go to Neurology.org for full disclosures.**

Received January 17, 2011. Accepted in final form October 7, 2011.

## REFERENCES

1. Wiklund I. Assessment of patient-reported outcomes in clinical trials: the example of health-related quality of life. *Fundam Clin Pharmacol* 2004;18:351-363.
2. Heffernan C, Jenkinson C. Measuring outcomes for neurological disorders: a review of disease-specific health status instruments for three degenerative neurological conditions. *Chronic Illn* 2005;1:131-142.
3. Epton J, Harris R, Jenkinson C. Quality of life in amyotrophic lateral sclerosis/motor neuron disease: a structured review. *Amyotroph Lateral Scler* 2009;10:15-26.
4. Leone MA, Beghi E, Righini C, Apolone G, Mosconi P. Epilepsy and quality of life in adults: a review of instruments. *Epilepsy Res* 2005;66:1-3.
5. Riazi A. Patient-reported outcome measures in multiple sclerosis. *Int MS J* 2006;13:92-99.
6. Den Oudsten BL, van Heck GL, De Vries J. The suitability of patient-based measures in the field of Parkinson's disease: a systematic review. *Mov Disord* 2007;22:1390-1401.
7. Nowinski CJ, Victorson D, Cavazos JE, Gershon R, Cella D. Neuro-QOL and the NIH Toolbox: implications for epilepsy. *Therapy* 2010;7:533-540.
8. Cella D, Nowinski C, Peterman A, et al. The neurology quality-of-life measurement initiative. *Arch Phys Med Rehabil* 2011;92(suppl):S28-S36.
9. Cella D, Victorson D, Nowinski C, Peterman A, Miller DM. The Neuro-QOL project: using multiple methods to develop a HRQOL measurement platform to be used in clinical research across neurological conditions. *Qual Life Res* 2006;14:1353.
10. Miller D, Nowinski C, Victorson D, Peterman A, Perez L. The Neuro-QOL project: establishing research priorities through qualitative research and consensus development. *Qual Life Res* 2005;14:2031.
11. Perez L, Huang J, Jansky L, et al. Using focus groups to inform the Neuro-QOL measurement tool: exploring patient-centered, health-related quality of life concepts across neurological conditions. *J Neurosci Nurs* 2007;39:342-353.
12. Lynch EB, Butt Z, Heinemann A, et al. A qualitative study of quality of life after stroke: the importance of social relationships. *J Rehabil Med* 2008;40:518-523.
13. Reise SP, Waller NG. Item response theory and clinical measurement. *Annu Rev Clin Psychol* 2009;5:27-48.
14. Rao D, Choi SW, Victorson D, et al. Measuring stigma across neurological conditions: the development of the stigma scale for chronic illness (SSCI). *Qual Life Res* 2009;18:585-595.
15. Gershon RC, Lai J-S, Bode R, et al. Quality of life item banks for adults with neurological disorders: item development and calibrations based upon clinical and general population testing. *Qual Life Res Epub* 2011 Aug 27.
16. Cella D, Gershon R, Lai JS, Choi S. The future of outcomes measurement: item banking, tailored short-forms, and computerized adaptive assessment. *Qual Life Res* 2007;16:133-141.
17. Chang CH. Patient-reported outcomes measurement and management with innovative methodologies and technologies. *Qual Life Res* 2007;16(suppl 1):157-166.
18. Cook KF, Choi SW, Crane PK, Deyo RA, Johnson KL, Amtmann D. Letting the CAT out of the bag: comparing computer adaptive tests and an 11-item short form of the Roland-Morris Disability Questionnaire. *Spine* 2008;33:1378-1383.
19. Hays R, Bode R, Rothrock N, Riley W, Cella D, Gershon R. The impact of next and back buttons on time to complete and measurement reliability in web-based surveys. *Qual Life Res* 2010;19:1181-1184.

## Thank you, Dr. John F. Kurtzke!

The *Neurology*<sup>®</sup> online archive has recently been updated to include the following issues:

- June 1968; 18 (6 Part 2):1-10
- May 1970; 20 (5 Part 2):1-59
- February 1988; 38 (2):309-316

The Editorial Office acknowledges Dr. John F. Kurtzke for his assistance in filling these gaps.