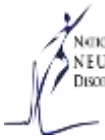
Neuro•QOL Quality of Life in Neurological Disorders

Measuring Quality of Life in Neurological Disorders

Final Report of the Neuro-QOL Study

September, 2010



FROM THE NEURO-QOL PRINCIPAL INVESTIGATOR

On behalf of the Neuro-QOL Executive Committee, co-investigators and collaborative partners, it is my great privilege to submit this Final Report of the Neuro-QOL study to the National Institutes of Neurological Disorders and Stroke. This document is the product of a highly collaborative and innovative effort involving numerous individuals and institutions, to whom we are incredibly grateful.

It is our hope that this work will raise the necessary interest and awareness of quality of life measurement in the neurology research community, and lead to a more profound understanding and appreciation of the daily realities and unique challenges faced by individuals with neurological disorders. We fully expect the Neuro-QOL tools that have been developed from this effort to serve as a foundation for continued QOL measurement programs and initiatives and look forward eagerly to helping to advance this mission and scientific knowledge base.

Sincerely,

David Cella, Ph.D.
Professor and Chair
Department of Medical Social Sciences
Northwestern University Feinberg School of Medicine

ACKNOWLEDGMENTS

The Neuro-QOL measurement system has been a collaborative effort of the National Institutes of Neurological Disorders and Stroke and several partnering institutions including Northwestern University, NorthShore University Healthcare, the University of Chicago, Cleveland Clinic Foundation, Children's Memorial Hospital of Chicago, Dartmouth Hitchcock Medical Center, the University of Pennsylvania, the University of Puerto Rico, the University of California – Davis Health System, the Rehabilitation Institute of Chicago, the University of Texas Health Science Center – San Antonio, the University of North Carolina – Charlotte and Westat.

In addition to these partnering institutions and individuals, the Neuro-QOL team wishes to acknowledge and extend special thanks to the following groups and individuals:

- The numerous scientists, clinicians, community advocates, and other professionals who generously gave of their time and expert knowledge. Without their participation, the Neuro-QOL tool would not have been possible.

- In particular, we acknowledge the participants in our Neuro-QOL Blue Ribbon Panel who met for a daylong consensus meeting to identify priority neurological conditions from which to base this tool:

Name	Area of Expertise	Institution
Jose Biller, MD	Stroke	Loyola University Medical Center
Bruce Dobkin, MD	SCI/TBI	UCLA School of Medicine
Pauline Filipek, MD	Autism – Pediatric	University of California – Irvine
Jacob Fox, MD	AD/Dementia	Rush University Medical Center
Deborah Gaebler-Spira, MD	Cerebral Palsy	Rehabilitation Institute of Chicago
Christopher Goetz, MD	Parkinson's Disease	Rush University Medical Center
Gregory Holmes, MD	Epilepsy	Dartmouth-Hitchcock Medical Center
Irene Katzan, MD, MS	Stroke	Cleveland Clinic Foundation
Nabih Ramadan, MD	Headache	Rosalind Franklin University of Science & Medicine
Anthony Reder, MD	MS	University of Chicago
Robert Sufit, MD	ALS	Northwestern University Feinberg School of Medicine
Jerry Sweet, PhD	Neuropsychology	NorthShore University HealthSystem
Elaine Wyllie, MD	Pediatric Epilepsy	Cleveland Clinic Foundation

- The many patients and caregivers who participated in focus groups, individual interviews and who completed the Neuro-QOL items during development and validation testing.

- The NINDS, under the leadership of the Neuro-QOL Project Officer Dr. Claudia Moy, who provided ongoing guidance, scientific support and encouragement throughout the contract period.

TABLE OF CONTENTS

1. EXECUTIVE SUMMARY……6

2. INTRODUCTION……7

3. DEVELOPMENT OF THE NEURO-QOL TOOL……8
   - 3.1. Identifying Target Neurological Conditions……8
   - 3.2. Establishing Criteria for Acceptance from End Users……10
   - 3.3. Identifying Health Related Quality of Life Domains……11
     - 3.3.1. Literature Review……11
     - 3.3.2. Expert Interviews……12
     - 3.3.3. Patient and Caregiver Focus Groups……14
     - 3.3.4. Keyword Literature Search……15
     - 3.3.5. Development of a Preliminary Domain Framework……15
   - 3.4. Selection of Generic Domains: Overall Method……16
   - 3.5. Selection of Generic Sub-Domains: Overall Method……18
     - 3.5.1. Expert Interview II Domain Elaborations……18
     - 3.5.2. Focus Group Findings……20
   - 3.6. Selection of Targeted Domains: Overall Method……22
   - 3.7. Neuro-QOL "Develop but not Test" Item Pools……23
   - 3.8. Final Adult and Pediatric Domain Frameworks……24
   - 3.9. Selection and Development of Items……26
     - 3.9.1. Identification of Data Sources……26
     - 3.9.2. Organization and Classification of Items……28
       - 3.9.2.1. Development of an Item Library……28
       - 3.9.2.2. Binning……28
       - 3.9.2.3. Winnowing……29
       - 3.9.2.4. Qualitative Item Review……29
       - 3.9.2.5. Criteria for Evaluating Items……31
     - 3.9.3. Spanish Translation……32
     - 3.9.4. Cognitive Interviewing with Patients and Caregivers……36
     - 3.9.5. Field Testing Ready Item Pools for Adults and Children……36

4. CALIBRATION AND VALIDATION TESTING……37
   - 4.1. Overview……37
   - 4.2. Wave 1a Online Clinical Sample……38
   - 4.3. Wave Ib General Population Sample……38
   - 4.4. Wave II Clinical Samples……38
   - 4.5. Analysis Plan and Item Calibrations……39
   - 4.6. Development of Short Forms……41
   - 4.7. Wave II Clinical Validation Study……41
     - 4.7.1. Methods……42
       - 4.7.1.1. Participating Sites……42
       - 4.7.1.2. Site Procedures……42
       - 4.7.1.3. Inclusion/Exclusion Criteria……43
       - 4.7.1.4. Recruitment and Testing……44
       - 4.7.1.5. Overview of Measures and Administration Schedule……44
       - 4.7.1.6. Cross Disease Measures……44
       - 4.7.1.7. Disease Specific Measures……48
       - 4.7.1.8. Statistical Analyses……50
   - 4.8. Wave II Results……51
     - 4.8.1. Results – Stroke Sample……51
       - 4.8.1.1. Sample Characteristics……51
       - 4.8.1.2. Reliability……52
       - 4.8.1.3. Validity……52

4.8.1.4. Known Groups Validity…… 55
4.8.1.5. Responsiveness…… 55
4.8.1.6. Conclusions……56
4.8.2.Results – Amyotrophic Lateral Sclerosis Sample…… 56
4.8.2.1. Sample Characteristics……56
4.8.2.2. Reliability……56
4.8.2.3. Validity…… 57
4.8.2.4. Known Groups Validity…… 59
4.8.2.5. Responsiveness…… 59
4.8.2.6. Conclusions……60
4.8.3.Results – Multiple Sclerosis Sample……60
4.8.3.1. Sample Characteristics……60
4.8.3.2. Reliability……60
4.8.3.3. Validity…… 63
4.8.3.4. Known Groups Validity…… 64
4.8.3.5. Responsiveness…… 64
4.8.3.6. Conclusions……64
4.8.4.Results – Parkinson's disease Sample…… 65
4.8.4.1. Sample Characteristics……65
4.8.4.2. Reliability……65
4.8.4.3. Validity…… 65
4.8.4.4. Known Groups Validity…… 67
4.8.4.5. Responsiveness…… 67
4.8.4.6. Conclusions……68
4.8.5.Results – Adult Epilepsy Sample…… 68
4.8.5.1. Sample Characteristics……68
4.8.5.2. Reliability……68
4.8.5.3. Validity…… 69
4.8.5.4. Known Groups Validity…… 71
4.8.5.5. Responsiveness…… 71
4.8.5.6. Conclusions……72
4.8.6.Results – Pediatric Epilepsy Sample…… 73
4.8.6.1. Sample Characteristics……73
4.8.6.2. Reliability……73
4.8.6.3. Validity…… 73
4.8.6.4. Known Groups Validity…… 75
4.8.6.5. Responsiveness…… 75
4.8.6.6. Conclusions……76
4.8.7.Results – Muscular Dystrophy…… 76
4.8.7.1. Sample Characteristics……76
4.8.7.2. Reliability……76
4.8.7.3. Validity…… 77
4.8.7.4. Known Groups Validity…… 79
4.8.7.5. Responsiveness…… 79
4.8.7.6. Conclusions……79
4.8.8.Overall Conclusions……79
5. CONCLUSION……80
6. REFERENCES…… 81

1. EXECUTIVE SUMMARY

Neuro-QOL is an NINDS sponsored initiative that developed and validated clinically relevant and psychometrically robust health-related quality of life (HRQL) assessment tools for adults and children with neurological disorders. The funding period for the Neuro-QOL project was from September 30, 2004 - September 30th, 2010.
Since many of the traditional clinical or functional measures of disease status have not adequately represented the full scope of the impact on an individual of chronic neurological disorders and their treatments, Neuro-QOL was conducted to accomplish this task. The Neuro-QOL measurement system is responsive to the needs of researchers in a variety of neurological disorders and settings and can facilitate comparisons of data across clinical trials in different diseases. Through an extensive patient centered conceptual and item development phase followed by two separate waves of large scale testing and analysis with US General Population and clinical samples, the Neuro-QOL measurement system includes precise and responsive patient reported outcome tools that span generic concepts such as physical, social, emotional and cognitive functioning, with additional disease relevant self report tools that reflect important symptoms and issues known to be important to patients and caregivers with neurological disorders.

2. INTRODUCTION

Neurologic disorders and their treatments can affect a wide array of physical, mental and social functioning. Since many neurologic conditions are chronic and incurable, treatment tends to focus on symptom management, limiting the extent of disability, and preventing disease progression. In short, treatment typically aims to improve the quality of patients' lives by limiting disease impact. Traditional clinical and functional measures of disease status do not represent the full impact of these conditions and their treatments. Multi-dimensional patient-reported outcome measures, such as health-related quality of life (HRQL) instruments that assess social, physical, and mental well-being, would be of greater value in this regard, particularly in clinical trials where differences in clinical measurements may or may not be significant. While there has been an increase in the development of neurology-specific HRQL tools and the incorporation of existing HRQL measures into neurology clinical trials, some of these questionnaires have questionable validity or may be difficult to interpret in this setting. There is little consensus on best tools and approaches, hindering the ability to make cross-disease and cross-study comparisons of relative disease burden, benefits of different treatments or other factors.

In order to address these issues, the National Institute of Neurological Disorders and Stroke (NINDS) sponsored Neuro-QOL, a 5-year, multi-site project to develop a bilingual (English/Spanish), clinically relevant and psychometrically robust HRQL measurement system for major neurologic conditions. Neuro-QOL has developed IRT-based patient reported outcomes of functioning across social, mental and physical well-being, paving the way to efficient, flexible and responsive assessment. This 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 HRQL issues. To accomplish this, Neuro-QOL developed and tested item banks, or finite sets of questions, assessing common concepts that cut across virtually all selected diseases. Added to these generic item banks are separate sets of unique, targeted scales evaluating symptoms, concerns or issues that are relevant only to a subset of diseases or treatments. Using modern psychometric methods, items in the banks are being used to construct computer adaptive tests (CATs) and short forms that are brief enough to be used in a variety of settings. The primary end users of this measurement system will be clinical trialists and other clinical neurology researchers; however it will also be appropriate for clinical practice.

3. DEVELOPMENT OF THE NEURO-QOL TOOL

3.1 Identifying Target Neurological Conditions. We identified 5 adult and 2 pediatric target neurological conditions through: a) critically examining the neurology research literature; b) interviewing numerous professionals involved in neurology clinical trials research and practice; and c) achieving consensus from a panel of renowned experts in neurology during a day-long face-to-face meeting. Methods undertaken were comprehensive and iterative and involved ongoing feedback to and from the NINDS for guidance and consultation. To this end, we completed a summary of disease prevalence and incidence information on 24 neurological conditions. In addition, we completed an extensive literature review of these conditions which focused on existing HRQL measures, issues regarding measurement for chronic neurological conditions, and disease-specific characteristics. This literature review summarized major neurological disorders and their concomitant impact on HRQL. It began with an overview of HRQL measurement issues in neurological disorders followed by the criteria with which we selected the featured disorders in the paper. Major neurological conditions were included, beginning with those typical to childhood onset followed by those most common in adults and advancing age.

In addition, we conducted interviews with 44 experts in neurological disorders and/or HRQL that asked for their opinion on: which neurological conditions to focus on and why; important HRQL domains for neurological disorders and what criteria an HRQL measure for neurological disorders should have. Information on target neurological conditions from the Expert Interviews was analyzed, summarized, and distributed to relevant Neuro-QOL affiliates. These 44 experts had practiced a median of 20 years (range 8 – 44), were primarily male (70%) and came from the following professions: Neurology (57%), Physiatry (14%), Health/Rehab Psychology (7%), Neuropsychology (7%), Nursing (4%) and Other (11%). Thirty-one (70%) experts saw only adult patients, 16% saw only pediatric patients, and 14% saw both. Most (89%) had been an investigator in a clinical trial. Thirty-two (73%) had used HRQL scales in clinical trials and thirty-three (75%) assess HRQL in clinical practice. The expert interviewees were asked to identify the five neurological disorders for which they felt it was most important to measure HRQL. They were not asked to specify whether they were nominating adult or pediatric diseases. Table 1 lists the disorders and the number of nominations they received from the expert interviewees.

Table 1. Disorders Selected by Expert Interviewees (N= 44)

Stroke	35	SCI	12
Multiple Sclerosis	33	ALS	10
Parkinson's	27	CNS tumor	7
Epilepsy/seizure	24	Pain disorders	5
AD/Dementias	17	Neurodegenerative	5
TBI	16	Sleep disorders	4
Migraine	13	Neuropathies	4

The NINDS Neuro-QOL Consensus Meeting was held in Chicago on March 15th, 2005 for the purpose of selecting five adult and two pediatric neurological disorders to be the focus of subsequent HRQL measure development activities. See Table 2 for the consensus panel attendees and their respective areas of expertise and institutions.

Table 2. Neuro-QOL Consensus Panel of Experts in Neurological Disorders

Name	**Area of Expertise**	**Institution**
Jose Biller, MD	Stroke	Loyola University Medical Center
Bruce Dobkin, MD	SCI/TBI	UCLA School of Medicine
Pauline Filipek, MD	Autism – Pediatric	University of California – Irvine
Jacob Fox, MD	AD/Dementia	Rush University Medical Center
Deborah Gaebler-Spira, MD	Cerebral Palsy	Rehabilitation Institute of Chicago
Christopher Goetz, MD	Parkinson's Disease	Rush University Medical Center
Gregory Holmes, MD	Epilepsy–Adult & Pediatric	Dartmouth-Hitchcock Medical Center
Irene Katzan, MD, MS	Stroke	Cleveland Clinic Foundation
Nabih Ramadan, MD	Headache	Rosalind Franklin University
Anthony Reder, MD	MS	University of Chicago
Robert Sufit, MD	ALS	Northwestern University
Jerry Sweet, PhD	Adult Neuropsychology	Evanston Northwestern Healthcare
Elaine Wyllie, MD	Pediatric Epilepsy	Cleveland Clinic Foundation

Prior to selecting the target diseases, the consensus panel agreed to use the following criteria when making their decision:

1. Prevalence of the disease/disorder
2. Magnitude of the disease's impact on the individual
3. The existence of promising current or new treatments on the horizon
4. Multiple domains affected
5. Chronic nature of the disease/possibility of seeing HRQL change

In contrast to expert interviewees, consensus panel members were asked to select five adult and two pediatric diseases separately. The results of this process (with number of votes) are presented in Tables 3 and 4.

Table 3. Adult Diseases Selected by Consensus Panel

Stroke	13	Migraine Headache	7
Multiple Sclerosis	12	Spinal Cord Injury	6
Parkinson's Disease	11	Epilepsy	5
Alzheimer's Disease	10	Traumatic Brain Injury	1

Table 4. Pediatric Diseases Selected by Consensus Panel

Epilepsy	9
Muscular Dystrophies	7
Cerebral Palsy	4

In a recommendations report to the NINDS, we suggested that the following three conditions be included for adult disorders: Stroke, Multiple Sclerosis and Parkinson's disease. These conditions were the top three conditions selected by the expert panel's final vote. They were also the top three conditions nominated by the 44 surveyed experts, and they represent prevalent neurological conditions with extensive impact upon an individual patient's quality of life. We also recommended that further consideration be given to the following conditions, based on their ranking from experts surveyed and consensus panel experts: Epilepsy/Seizure Disorders, Alzheimer's Disease/Dementia, Traumatic Brain Injury, Spinal Cord Injury and Migraine Headache. As a result of these recommendations and NINDS feedback, the Executive Committee finalized this list of adult and pediatric conditions (see Table 5), which included: 1) Adult Conditions - Stroke, Multiple Sclerosis, Parkinson's Disease, Epilepsy and Neuromuscular Disorders (which has been defined as Amyotrophic Lateral Sclerosis and Myasthenia Gravis); 2) Pediatric Conditions – Epilepsy and Muscular Dystrophies

Table 5. Final Recommendations

Adult Conditions	
	Stroke
	Multiple Sclerosis
	Parkinson's disease
	Epilepsy
	Neuromuscular Disorders (Amyotrophic Lateral Sclerosis and Myasthenia Gravis)
Pediatric Conditions	
	Epilepsy
	Muscular dystrophies

3.2 Establishing Criteria for Acceptance from End Users. As the primary goal of this project is to develop an HRQL measure for widespread use in neurology clinical trials and clinical research, a key first task was to identify criteria for the acceptance of HRQL measures in these communities. The data obtained is intended to inform the subsequent stages of the project to maximize the probability that the final HRQL instrument will meet with widespread acceptance among the likely users. To accomplish this, we set forth to obtain data on the knowledge of, attitudes toward, and use of HRQL measures and information by physicians and other professionals who participate in clinical neurology research, including clinical trials in particular. Using an adapted version of the MD-QOL, potential respondents were identified from several sources including our clinical testing sites and consultants, NINDS reviewers and grantees, the American Academy of Neurology and the American Congress of Rehabilitation Medicine. We sent a detailed letter to neurology professionals requesting their assistance and explaining our interest in developing comprehensive item banks that will be used to create generic and targeted questionnaires for clinical trials across major neurological conditions. Each respondent who expressed interest in participating was sent a web address to a password- protected, confidential online survey that contained the MD-QOL Neurology as well as demographic and qualitative questions. A total of 103

Neurology surveys were received, of which 89 had complete data. Respondents were asked to list characteristics that a HRQL questionnaire must have in order for it to be useful to the neurology clinical research community. Grouping of similar characteristics into more general categories revealed several important criteria that an HRQL measure should possess. The first criterion was satisfactory psychometric properties (38% of all responses). Respondents emphasized that the final measure should have adequate population-specific validity, reliability, responsiveness, specificity and precision and sensitivity to change. The second most frequently listed criterion was ease of administration and use (28%). The measure should be minimally burdensome for the patient and clinician, able to be completed by people at various levels of disability and literacy, include alternative modes of administration (e.g., phone, computer), use unambiguous questions and be brief (less than 10-15 minutes). The third criterion related to the content of the measure (12%) and the need to include the diversity of symptoms and HRQL domains impacted by neurological disorders Finally, the measure should be clinically relevant and have direct application to patient care and applied outcomes (11%). The remaining responses addressed the need for the measure to be objective, have good normative data and be patient-centered.

3.3 Identifying Health Related Quality of Life Domains. We identified domains through multiple methods and data sources. This included a comprehensive review of the literature and literature search, expert interviews and surveys and patient and caregiver focus groups.

3.3.1 Literature Review. Initially, we identified conditions through completing an extensive Medline literature review of 24 neurological conditions using key words such as HRQL, neurological disorders, measurement issues, as well as disease-specific characteristics, from 1996 to the present. This literature review summarized major neurological disorders and their concomitant impact on HRQL. Beginning with those typical to childhood onset followed by those most common in adults and advancing age, major neurological conditions included:

Table 6. Major Neurological Conditions in Literature Review

CONDITIONS WITH LIKELY ONSET IN CHILDHOOD AND ADOLESCENCE

- DUCHENNE MUSCULAR DYSTROPHY
- CEREBRAL PALSY
- AUTISM
- EPILEPSY
- PRIMARY BRAIN TUMORS
- ATTENTION DEFICIT HYPERACTIVITY DISORDER

CONDITIONS WITH LIKELY ONSET IN ADOLESCENCE AND EARLY ADULTHOOD

- NARCOLEPSY
- TRAUMATIC BRAIN INJURY
- SPINAL CORD INJURY

CONDITIONS WITH LIKELY ONSET IN EARLY AND MIDDLE ADULTHOOD

EPILEPSY
MULTIPLE SCLEROSIS
MIGRAINE HEADACHE
COMPLEX REGIONAL PAIN SYNDROME

CONDITIONS WITH LIKELY ONSET IN MIDDLE AND LATE ADULTHOOD

STROKE
PARKINSON'S DISEASE
HUNTINGTON'S DISEASE
AMYOTROPHIC LATERAL SCLEROSIS
PRIMARY BRAIN TUMORS
ALZHEIMER'S DISEASE
CHRONIC PAIN SYNDROME
DIABETIC PERIPHERAL NEUROPATHY
IDIOPATHIC PERIPHERAL NEUROPATHY
HEREDITARY MOTOR-SENSORY NEUROPATHY

From this review, our initial list of tentative domains included:

- Emotional distress
- Cognitive function
- Social function
- Physical function
- Mobility
- Fatigue
- Pain
- Communication/language
- Sexual function
- Sleep disturbance
- Independence
- Role Participation
- Sensory Impairment
- Global HRQL
- Fine Motor/ADLs
- Treatment Effects
- Stigma
- Bowel/bladder function

3.3.2 Expert Interviews. We completed two waves of expert interviews and an online request for information to identify the most important HRQL domains affected by neurological conditions. In our first wave of interviews (n=44), among other tasks, we asked respondents to identify domains or areas of QOL that are affected by neurological disorders and their treatments. Since these interviews occurred before we had selected our seven Neuro-QOL conditions (Task II), there was a low prevalence of expertise in certain conditions, such as pediatric epilepsy, ALS or muscular dystrophy. Experts were informed that their responses could include important symptoms (e.g., pain), areas of function (e.g., mobility), or anything else that was deemed important to consider when thinking of the QOL of people with neurological disorders. Experts were first asked to list all the domains they believed would be important to cover in a QOL questionnaire that could be given to patients with any neurological disorder (i.e., general neurological patients). After that, they were asked to list domains that might be important in one of the disorders they named previously, but that wasn't necessarily common to all disorders.

Table 7. Experts from Individual Interviews and Online Request for Information

	Interview I (n=44)	Interview II (n=63)	Online Request for Information (n=89)
Years in Practice (median)	20	21	18
Male	70%	70%	56%
Profession			
Neurology	57%	43%	63%
Physiatry	14%	18%	5%
Health/Rehab Psychology	7%	9%	2%
Neuropsychology	7%	7%	5%
Nursing	4%	2%	9%
Other	11%	21%	5%
Adult patients only	70%	78%	65%
Pediatric patients only	16%	8%	16%
Both	14%	14%	17%
Investigator in a clinical trial	89%	89%	NA
Use HRQL scales in research	73%	56%	NA
Use HRQL scales in practice	75%	29%	NA

We also conducted an online request for information from experts in neurology (n=89) and asked them to think about their patients who participate in clinical research and list the five neurological disorders for which they think it is most important to measure HRQL. For each disorder, they were asked to identify the three areas of well-being or function (HRQL Domain) that they believe should be included in a HRQL questionnaire for that disease. These interviews confirmed domains that had been identified from the literature review and also revealed the following new areas:

- Behavior/Personality Change
- Driving
- Memory
- Attention
- Executive function
- Aggression/irritability
- Psychotic symptoms
- Meaning and spirituality
- Mastery and control

Finally, we completed a second wave of expert interviews (n=63) and asked them to think about all of the different ways that their subspecialty disease(s) in particular could affect patients' quality of life and to list all domains they believe would be important to provide item coverage in a HRQL questionnaire. These interviews provided greater depth and elaboration of content for given domains, as seen in the example in Table 8.

Table 8. Example of Domain Elaboration from Expert Interviews II

DISEASE	DOMAIN	SUB-DOMAIN	%
	Social Function	Accessibility	4.2%
	Physical Function	ADLs	11.5%
	Personality/Behavioral Change	Aggression	6.3%
	Emotional Distress	Alexithymia	3.4%
	Personality/Behavioral Change	Anger	3.1%
Stroke	Emotional Distress	Anxiety	13.8%
	Personality/Behavioral Change	Apathy	3.1%
	Perceived Cognitive Function	Apraxia	3.4%
	Communication/Language Difficulty	Articulation	2.4%
	Perceived Cognitive Function	Attention/concentration	17.2%

3.3.3 Patient and Caregiver Focus Groups (N=83). We conducted 11 focus groups with patients and caregivers (seven with patients (n=64); four with caregivers (n=19)) to assess the impact of neurological conditions on quality of life domains. We began with broad questions, allowing participants to free-list responses on their definition of quality of life. We then progressed to questions regarding specific quality of life domains, such as physical function, emotional function, social aspects, and treatment effects that have been shown to be relevant in the literature. Focus groups with caregivers of Alzheimer's Disease, stroke, and pediatric epilepsy patients were conducted as these patients may be unable to reliably report their subjective perceptions of HRQL due to cognitive impairment or age. Participants' age in the adult focus groups ranged from 25 – 84 years (mean = 52.1, median = 55) and the age range for the pediatric group was 14 - 20 years (mean = 16.38, median = 16). For the adult caregiver groups, the age range of their care recipient was 55 – 82 years (mean = 60.39, median = 65). For the pediatric caregiver group, the age range of their care recipient was 11 -18 years (mean = 11.29, median =12). Mean and median ages for participants of patient groups are as follows: MS (mean =49.7, median = 46), ALS (mean = 60.67, median = 65), epilepsy (mean = 39.75, median= 40), Parkinson's Disease (mean = 63.7, median = 61.5), Alzheimer's Disease (mean =72.14, median 75), and stroke (mean = 54.67, median = 57), pediatric epilepsy (mean =14.14, median = 16). Mean and median ages for the care recipients of the caregiver groups are as follows: Alzheimer's Disease (mean = 62.44, median = 65) and stroke (mean = 65.63, median = 63). For the care recipients of the pediatric epilepsy caregiver group, the mean was 13.2 and the median was 12 years. Please see Table 9 for an example of domain code percentages that were applied. For a more extensive review of focus group findings, please see Perez et al. [1]

Table 9. Example of Frequency of Emotion Domain Codes from Focus Groups

	ALS%	Pediatric Epilepsy%	Adult Epilepsy%	Multiple Sclerosis%	Parkinson's Disease%	Stroke%
Denial	0	0	4.3	2.5	0.9	0.5
Frustration/loss	0	0	0	0.5	0.9	3.1
Frustration/cognitive	0	0	0	1.2	.0.	3.1
Depression	0	1.6	4.3	0.7	0	1.8
Isolation	0	0	4.3	0.7	0	0
Fear	0	0	0	0	0.9	0
Embarrassment	1	0	4.3	1.7	2.7	7.6
Grieving	1	0	0	0.7	0	0
Anger	0	0	2.1	0.3	2.7	4.5
Guilt	3.1	0	2.1	0	0.9	0.9
Mood changes	0	0	0	0	0	2.7
Emotional Function Total	**5.1**	**1.6**	**21.3**	**8.4**	**9**	**24.1**

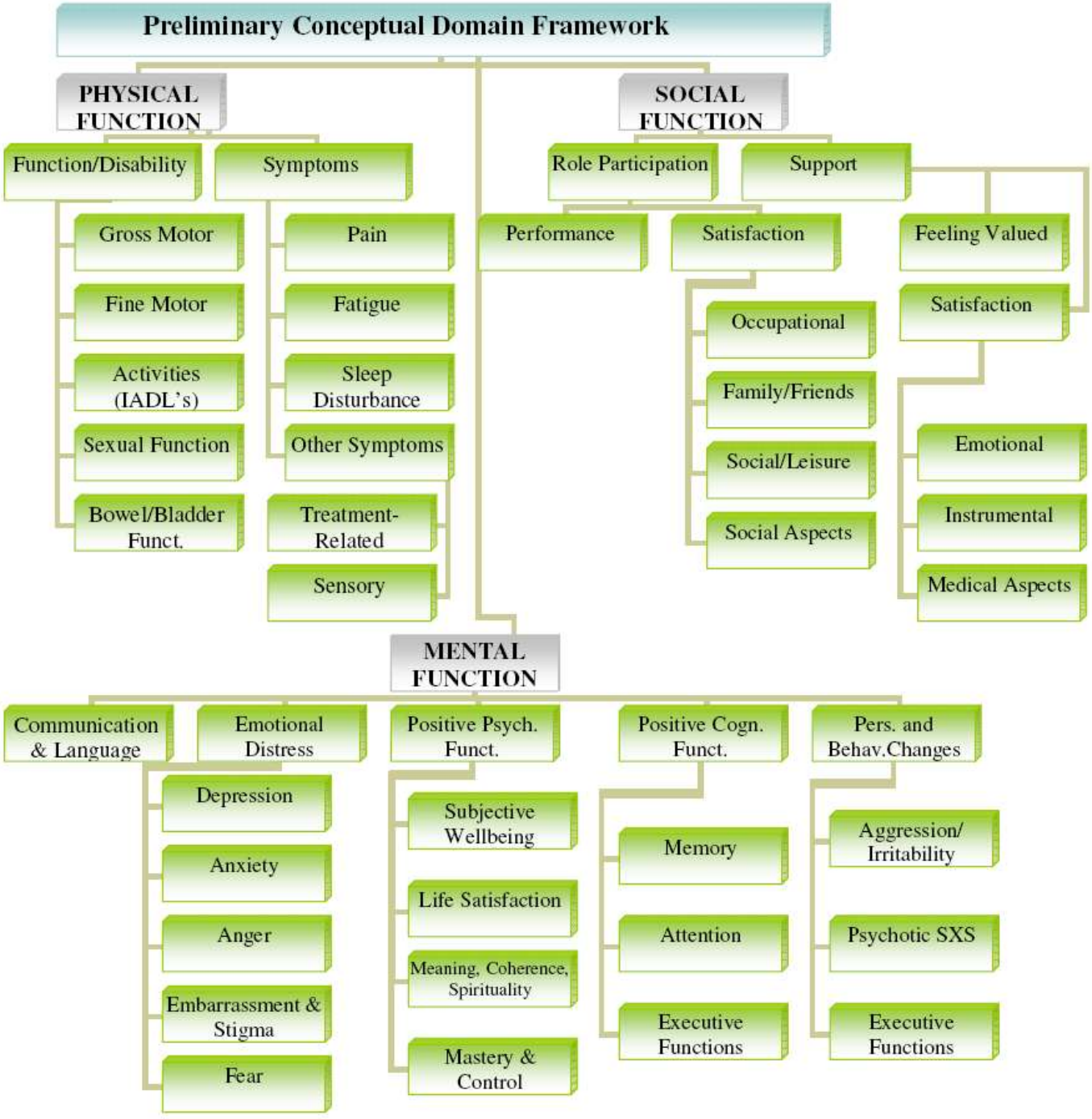
3.3.4 Keyword Literature Search. Because new domains arose from these different sources, we also conducted a comprehensive keyword literature search (from 1996 to present) via OVID search engine using previous and newly identified domains and Neuro-QOL diseases to best estimate the number of published studies in a given area. We felt that these totals, albeit approximate, would provide us with an overall quantification of how important certain domains were within different neurological conditions.

Table 10. Example of Keyword Literature Search

		ALS	**MS**	**Ped Ep.**	**Adult Ep.**	**PD**	**Stroke**	**MD, Duchenne**
	raw counts	*2,851*	*9,709*	*8,972*	*6,001*	*11,591*	*20,352*	*776*
PHYSICAL								
Fine/Gross motor	*41,325*	47	133	140	109	889	705	13
Bowel/Bladder	*28,783*	9	114	38	16	76	79	4
Sexual Function	*8,808*	0	47	9	13	47	10	0
ADL	*16,803*	30	197	38	30	277	677	14
Sensory	*100,994*	28	321	257	264	334	839	7
Deglutition	*1,809*	3	5	2	4	18	64	0
Fatigue	*4,755*	8	195	17	9	28	25	1
Pain	*54,819*	14	158	220	197	63	387	3
Sleep	*11,587*	12	12	153	59	109	49	1

***3.3.5 Development of a Preliminary Domain Framework*.** As a means of organizing the multiple domains that arose from the aforementioned data sources, we constructed a preliminary domain framework to guide in subsequent decision making, as seen in Figure 1. Our next task would be to identify which domains should be selected for generic item bank and targeted scale development.

Figure 1. Preliminary Conceptual Domain Framework



3.4 *Selection of Generic Domains: Overall Method.* First, we identified disease chairs from the Neuro-QOL Executive Committee and co-investigator panel. Each chair was assigned a disease from the seven Neuro-QOL conditions, which included stroke, adult epilepsy, ALS, Parkinson's disease, multiple sclerosis, muscular dystrophy (pediatric), and pediatric epilepsy. Each chair was charged with reviewing the aforementioned data sources and extracting the most relevant domains for item bank consideration. Methods undertaken were comprehensive and iterative and involved ongoing feedback to and from the NINDS for guidance and consultation.

Each source of data was analyzed using largely qualitative approaches. Although each disease chair allowed some degree of individual judgment in making domain selection decisions, this process primarily entailed identifying and coding content derived from context specific units of meaning. These codes were converted into percentages, which were calculated as the number of times a particular theme or code was applied over the total number of all codes applied from each data source. For example, using this extraction approach it was possible for us to understand how frequently physical function was mentioned in ALS, within the context of all other domains that were mentioned for ALS. This permitted us to better understand the occurrence (and by association, importance) of certain domains either across all conditions or as a unique aspect of one disease. Constant comparison to the literature and other sources of informant data was applied and enhanced the data collection process in an iterative manner.

Within each disease, domain percentages were calculated and recorded on a chart that was populated by information obtained from five sources as mentioned above: Expert Interview I, Expert Online Request for Information, Expert Interview II, Focus Group and Keyword Literature Search. For the two expert interviews and expert survey, we made a decision to only include the open-ended, spontaneously generated expert responses (vs. information experts suggested only after being asked to elaborate on a specific domain we provided them). We felt that by only using the domain information that was a part of their "brainstorming" process, we would capture the most important areas without inadvertently biasing their thought process with domains we felt were important or had identified from other sources. If a domain met a designated cutoff criterion across all five data sources, it received a sum of "5"; if it met the cutoff across four data sources, it received a sum of "4", and so on. These 0-5 counts were then compared across diseases. If a domain was counted as ≥3 on at least 50% of the diseases (e.g., 4/7 diseases) it was considered to be a generic concept. Targeted domains were those that summed ≥2 in at least one domain, but were not necessarily prevalent across the majority of diseases. In the event that certain disease specific domains "tied" either within or between conditions, we consulted our expert panel for their input. See Table 11 for generic and targeted domains.

Table 11. Domains and Importance Counts across Diseases

	EPILEPSY ADULT	MS	STROKE	PD	ALS	EPILEPSY PED	MD	Generic/ Targeted
Physical	2	5	5	5	5	2	4	Generic
Cognitive	4	3	4	5	2	3	2	Generic
Emotional	4	4	3	4	3	2	2	Generic
Social	4	4	4	4	5	4	4	Generic
Communication	2	1	2	2	3	1	1	Targeted
Fatigue	1	4	---	1	---	1	2	Targeted
Pain	2	1	2	1	2	1	2	Targeted
Treatment Effect	2	2	1	4	1	2	1	Targeted
Bowel & Bladder	---	2	---	1	---	---	1	Targeted
Independence	1	1	2	2	3	2	3	Targeted
Stigma	2	1	1	2	---	3	---	Targeted
Personality/ Behavior Change	1	1	1	1	1	1	2	Targeted
Positive Psychological Function	---	2	2	---	4	2	1	Targeted
Sensory SXS	1	1	1	1		1	1	NA
Sexual	1	1	1	1	1	---	1	NA

NOTE: Number in cell indicates the number of sources (5 = highest) that indicated the domain was of importance for the disease; Generic Concept = rating ≥3 in 50% of diseases; Targeted = ≥2 in less than 5 diseases.

After reviewing the findings of this comprehensive identification and selection process, the generic domains that were chosen were: Physical, Social, Emotional and Cognitive Function.

3.5 Selection of <u>Generic Sub-Domains</u>: Overall Method. Next, we identified domain co-chairs from the Neuro-QOL Executive Committee and co-investigator panel. Each co-chair was assigned a domain from the four generic domains previously selected and one pair was assigned the targeted domains. Each dyad was charged with reviewing the aforementioned data sources and extracting the most relevant sub-domains for item bank consideration (the dyad responsible for targeted scale decisions selected which scales should be developed and tested vs. only developed). Similar to the previous process, methods were extensive and involved continuous guidance and input from the NINDS. Again, data were analyzed using largely qualitative approaches, which are described below.

3.5.1 Expert Interview II Domain Elaborations*:* During these interviews, experts (7 per Neuro-QOL condition) were asked to elaborate upon several identified domains that might be affected by the neurological condition of their sub-specialty. All responses were exhaustively coded and grouped according to their disease and domain by two outcomes researchers familiar with neurological disorders and qualitative methodologies (See Table 12 for an example of this scheme).

Table 12. Example of Expert Interview II Coding Scheme for ALS

DISEASE	DOMAIN	EXPERT RESPONSE	CODE
ALS	Bowel/Bladder	Might need catheter or bedside commode	ASSISTIVE
		Nothing about disease inhibits function	
		Might not make it to bathroom b/c of gait weakness.	MOBILITY
		Incontinence is infrequent until later on in illness	INCONTINENCE
		Difficult getting to bathroom	MOBILITY

Next, percentages were calculated for the total number of times a particular code was applied within a domain (See Table 13). This was done to crudely estimate which codes might carry additional importance for a particular domain within a disease based on how often they were discussed among experts. We anticipate using this information to appropriately portion certain types of items within a given scale.

Table 13. Frequency of Applied Codes for Bowel/Bladder Function in ALS Example

DISEASE	DOMAIN	CODE	FREQUENCY	PERCENTAGE
ALS	Bowel/Bladder	ASSISTIVE	1	6.3%
		CONSTIPATION	1	6.3%
		DIARRHEA	1	6.3%
		INCONTINENCE	1	6.3%
		INCONTINENCE BL	2	12.5%
		INCONTINENCE BOW	1	6.3%
		MISC	2	12.5%
		MOBILITY	2	12.5%
		RESTRICTION	2	12.5%
		TOILETING	2	12.5%
		URGENCY BL	1	6.3%

The total number of applied codes was tallied both across and within conditions. The number of applied codes across conditions was used to determine which diseases shared similar codes relative to one another as well as which codes were unique to a particular disorder (See Table 14, left hand column). Decisions regarding generic sub-domains were based on issues that cut across the majority of diseases. The number of applied codes within conditions was used to estimate roughly how many different issues were relevant for each disease (See Table 14, bottom row). Decisions regarding targeted scales were based on which diseases contained the greatest number of issues, relative to other diseases.

Table 14. Bowel & Bladder Example Codes Across Conditions

DOMAINS/SUB-DOMAINS	ALS	Epilepsy	MD	MS	PD	Ped Epilepsy	Stroke	# DISEASES BY CODE
Bowel & Bladder								
ACCESSIBILITY			X					1
ASSISTIVE	X							1
CONSTIPATION	X		X	X	X	X		5
DEPENDENCE/INDEPENDENCE							X	1
DIARRHEA	X							1
EMBARASSMENT-2NDARY		X		X		X		3
FREQ				X				1
FREQUENCY BL				X	X			2
GENERAL PROBLEMS			X	X	X		X	4
HESITANCY				X				1
INCONTINENCE	X		X	X	X	X	X	6
INCONTINENCE BL	X	X	X	X	X	X	X	7
INCONTINENCE BOW	X	X	X	X		X	X	6
INFECTION				X				1
LEAKAGE					X			1
MAINTENANCE				X				1
MEDICATION-RELATED					X			1
MISC	X		X					2
MOBILITY-RELATED	X		X		X			3
OTHER		X						1
RESTRICTION	X		X					2
RETENTION				X	X			2
SECONDARY COMPLICATIONS					X			1
SPASMS					X			1
TOILETING	X		X					2
TROUBLE URINATING			X					1
URGENCY				X				1
URGENCY BL	X			X	X			3
URGENCY BOW				X				1
TOTAL # CODES BY DISEASE	**11**	**4**	**11**	**15**	**12**	**5**	**5**	

When reviewing this data to make generic sub-domain decisions, we referred to the total number of codes across diseases (see left hand column of Table 14) as a rough indicator to determine which sub-domains are highly predominant across the majority of Neuro-QOL diseases. When applicable, we gave greater importance to sub-domains that cut across at least 5-7 diseases. For example, in Table 14, if bowel and bladder were to be selected as a generic sub-domain for physical function, we would note that incontinence and constipation tend to be most prevalent across conditions compared to other types of bowel and bladder issues that were coded.

3.5.2 Focus Group Findings. Focus group participants were also facilitated through domain-elaboration discussions based on pre-identified domains as well as

newly elicited areas in which their respective neurological condition affects their quality of life. Responses underwent extensive open, axial and selective coding procedures by experts in qualitative methodologies (See Table 15). Percentages were calculated as the number of times a particular theme or code was applied over the number of all codes applied during for the focus group discussion

Table 15. Example of Focus Group Frequencies for Bowel and Bladder Function

Domain	Nodes	ALS %	PEP %	EP %	MS %	PD %	ST %
Bowel & Bladder Function	Bowel & Urinary Incontinence	0.0%	0.0%	0.0%	1.24%	0.0%	1.79%
	Frequent Urination	0.0%	0.0%	0.0%	2.48%	0.0%	0.89%

*Due to patient recruitment difficulties, no MD patient focus groups were conducted

When reviewing focus group data to make generic sub-domain decisions, we again referred to the frequencies and percentages of the total number of codes applied to domains across diseases (See Table 15). Because of the variability of responses, we simply used these percentages to provide confirmatory evidence to the expert interview data (e.g., were there any responses or not?). In some cases where focus group data were present regarding a particular sub-domain/condition association but expert interview data were not, we considered this as an area worth subsequent follow-up with other co-investigators and experts. Following this comprehensive review and selection process, the generic sub-domains we have chosen can be found in Tables 16 (adult) and 17 (pediatric).

Table 16. Adult Generic Sub-Domains for Development and Field Testing

Generic Domain	Sub-Domain I	Sub-Domain II	Sub-Domain III	Sub-Domain IV
Physical	Self-care/Upper Extremity	Mobility/ Ambulation	---	---
Social	Role Participation	Role Satisfaction	---	---
Emotion	Depression	Anxiety	Positive Psychological Function	Stigma
Cognitive	Perceived	Applied	---	---

Table 17. Pediatric Generic Sub-Domains for Development and Field Testing

Generic Domain	Sub-Domain I	Sub-Domain II	Sub-Domain III	Sub-Domain IV
Physical	Self-care/Upper Extremity	Mobility/ Ambulation	---	---
Social	---	---	---	---
Emotional	Emotional Health	Stigma	---	---

NOTE: Emotional Health covers areas of anxiety, depression, and worry

***3.6 Selection of Targeted Domains: Overall Method*.** Based on feedback from experts, as well as considering the complexity of issues surrounding these conditions, we decided to develop and field test select scales per condition, and also develop (but not field test) additional disease specific scales as indicated by the unique circumstances of each condition. To determine which scales would be field tested (vs. only developed), we summarized and examined data from three different sources: 1) Domain elaborations from our second round of expert interviews (N=63), findings from our focus groups (N= 11 groups), and expert panel ratings of their preferences for disease targeted scales (N= 25).

Following investigator meetings where generic domains were identified, we set forth to make preliminary decisions regarding which targeted scales should be developed, and for which disease(s). The results of these meetings led to the identification of a select number of candidate domains, which were presented to our panel of disease specific experts. Because the targeted domains presented to experts varied by disease (e.g., adult epilepsy experts were asked to rank fatigue, pain, bowel and bladder and stigma, while Parkinson's experts were asked to rank sleep, sexual function and personality/behavioral changes) it was not possible to rank each using the same denominator, but rather to examine each disease group individually. Using these expert rankings, along with Focus Group frequencies and the total number of coded targeted domain issues within each disease we identified our candidate targeted scales to develop and field test per disease, as well as additional targeted scales for development purposes only (see Table 18).

When reviewing this data to make targeted scale decisions, we referred to the total number of codes by disease as a rough indicator to determine which diseases are comparatively more affected by certain issues in a given domain. When applicable, we gave greater importance to domain-condition relationships when there was an approximate and sizeable difference between total codes among conditions. For example, in Table 11, MD, MS and PD all appear to have greater numbers of bowel and bladder issues that were coded, compared to adult/pediatric epilepsy, stroke and ALS.

When reviewing focus group data to make targeted scale decisions, we again referred to the frequencies and percentages of the total number of codes applied to domains across diseases. Because of the variability of responses, we simply used these percentages to provide confirmatory evidence to the expert interview data (e.g., where there any responses or not?). In some cases where focus group data were present regarding a particular domain-condition association but expert interview data were not, we considered this as an area worth subsequent follow-up with other co-investigators and experts. Following this comprehensive review and selection process, the targeted scales selected for development and field testing can be found in Table 18.

Table 18. Targeted Scales for Development and Field Testing

Condition	Develop and Field Test	Develop Only			
	1st choice	*2nd choice*	*3rd choice*	*4th choice*	*5th choice*
ALS	Fatigue/ Weakness	Bowel & Bladder	End of Life Concerns	---	---
Epilepsy	Fatigue/ Weakness	---	---	---	---
Muscular Dystrophy	Pain	Fatigue/Weakness	Bowel & Bladder	Personality and Behavioral Changes	---
Multiple Sclerosis	Fatigue/ Weakness	Bowel & Bladder	Sexual Function	Personality and Behavioral Changes	Sleep
Parkinson's Disease	Sleep	Personality and Behavioral Changes	Sexual Function	Bowel & Bladder	---
Pediatric Epilepsy	Fatigue Cognition	---	---	---	---
Stroke	Personality and Behavioral Changes	Sleep	Sexual Function	---	---

***3.7 Neuro-QOL "Develop but not Test" Item Pools*.** Based on findings from patient focus groups and expert interviews early in the contract period, it was determined that certain areas of functioning and quality of life were important to measure, yet were relatively lower in priority compared to other areas for which the Neuro-QOL system was developing and testing items banks. Because the scope and limitations of the contract resources and time period precluded the study team from including all possibly important domain areas, a decision was made to "develop but not test" item pools for the following areas: ADULTS: bowel function, bladder function, sexual function and end of life concerns; PEDIATRICS: bowel function, bladder function, emotional and behavioral dyscontrol. The first three domain areas were relevant to multiple diseases however the last domain was selected primarily for ALS. From the time this decision was made, item banks and scales of the same constructs have been developed and tested in related conditions or studies. For example, Bowel and Bladder/Urinary item banks have been developed and tested for men diagnosed with prostate cancer with funding from a grant from the American Cancer Society (Primary Contact: David Cella), of which Dr. Cella has agreed to make available to neurology researchers who wish to test and further validate in neurology samples. In

addition, Neuro-QOL investigators (Victorson, Cella, Heinemann) have assisted in the creation of additional Bowel and Bladder item banks through Neuro-QOL extension studies, SCI-QOL and TBI-QOL (Primary Contact: David Tulsky). These banks were created using Neuro-QOL methodologies and have been administered along several Neuro-QOL item banks with large clinical samples of individuals with SCI and TBI. These are also being made available for subsequent testing and use with different neurological disorders. The PROMIS Cancer Supplement (Primary Contact: Kevin Weinfurt) has developed an extensive sexual functioning items, as well as the SCI-QOL/TBI-QOL work, and therefore both are being suggested for further consideration and use in neurology research. Finally, Neuro-QOL investigators (Victorson, Peterman) have worked in collaboration with Dr. Noelle Carlozzi's NINDS-funded Huntington's Disease Quality of Life study (HD-HRQL) and have helped integrate Neuro-QOL ALS focus group information into her End of Life Concerns item bank so that it can be sufficiently useful to both ALS and HD populations. Dr. Carlozzi has given permission for these items to be used in subsequent neurology research studies as a part of Neuro-QOL efforts. Table 19 below provides the number of items from each respective project and item bank, as well as the primary contact person and email address. It is recommended that interested researchers contact these individuals before using any of these available items, as modifications may have occurred.

Table 19. Neuro-QOL "Develop but not Test" Item Pools

Study Name & Primary Contact	**P-QOL** David Cella d-cella@northwestern.edu	**CAPS (PROMIS)** Kevin Weinfurt kevin.weinfurt@duke.edu	**SCI-QOL /TBI-QOL** David Tulsky dtulsky@med.umich.edu	**HD-HROL** Noelle Carlozzi carlozzi@med.umich.edu	**Total**
Bowel Function	51	0	28	0	**79**
Urinary/ Bladder Function	42	0	16	0	**58**
Sexual Function	0	74	62	0	**136**
End of Life Concerns	0	0	0	61	**61**

* Since some of the items may have continued to undergo modification and testing, we advise interested researchers to contact the primary contact person prior to including any items in subsequent studies; ** Researchers interested in pediatric versions of bowel, bladder or emotional & behavioral dyscontrol will be able to base modifications on available adult items delivered under this contract

3.8 Final Adult and Pediatric Domain Frameworks. Once final domains were selected, the scientific team re-organized the Preliminary Conceptual Framework (Figure 1) to reflect the domain structure from which item banks and scales would be developed for further testing and validation. Figures 2 and 3 below highlight the adult and pediatric domain frameworks.

Figure 2. Adult Domain Framework for Item Banks and Scales

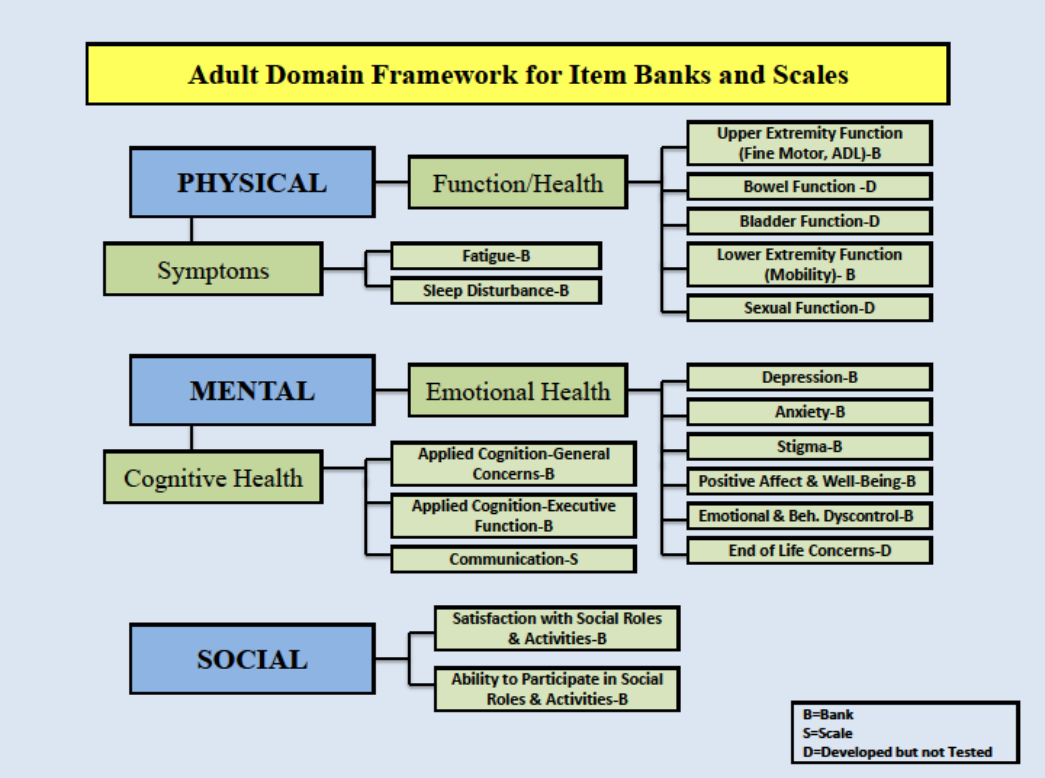
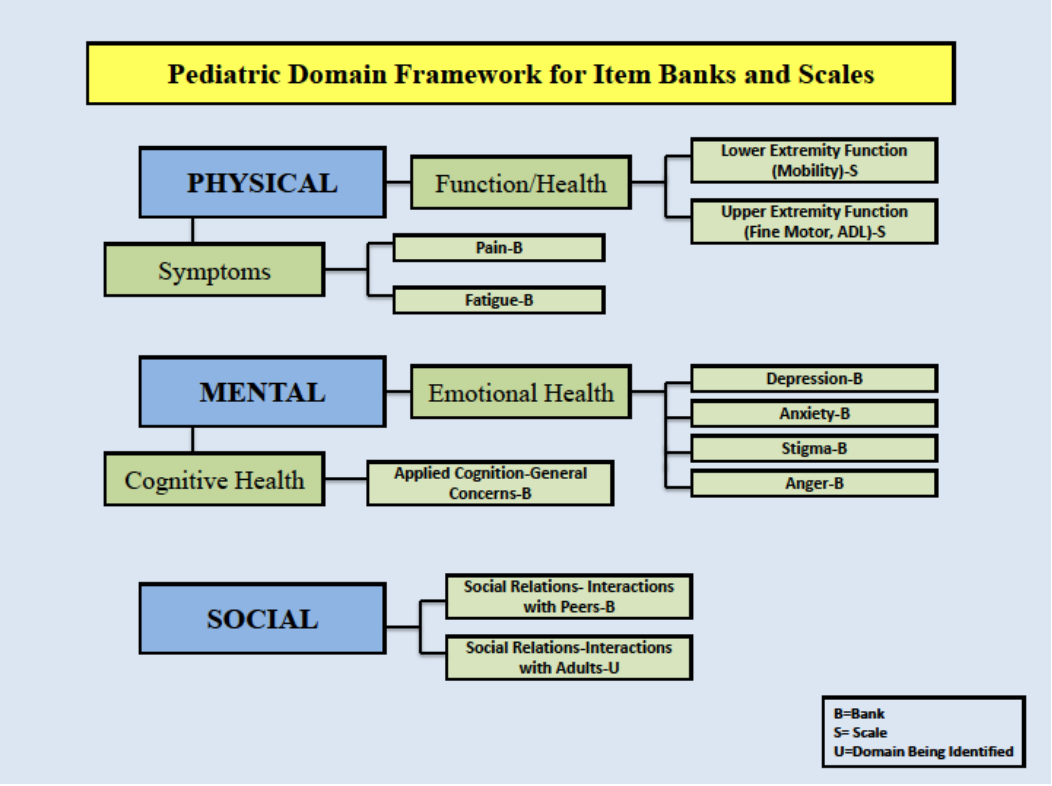


Figure 3. Pediatric Domain Framework for Item Banks and Scales



3.9 Selection and Development of Items. The process of identifying and selecting items to comprise Neuro-QOL's generic item banks and targeted scales was extensive and entailed multiple steps to assure the most comprehensive and content-relevant item pools.

3.9.1 Identification of Data Sources. Candidate items for the generic item banks and targeted scales were identified from CORE's existing item banking projects and affiliated studies, large external dataset analyses (for which we secured permission for analysis and use of scale items) and additional generic and disease-specific questionnaires that have been used in neurological conditions, with authors' permission.

For the generic domains and sub-domains selected for Neuro-QOL, which had corresponding item banks at CORE or affiliated projects, items were reviewed and entered into an organizing item library. Existing items and available large neurological datasets that contained item-level data on multiple QOL measures were also reviewed and analyzed through Rasch analysis methodology to help sharpen construct definitions of these QOL dimensions (see Table 20 for large datasets that were analyzed for these purposes). These data were evaluated by examining the content and dimensionality of the constituent items in these preliminary banks. Item location gaps (often found at the ceiling or floor) were determined, and expert review of content helped evaluate whether new content was needed. Candidate items from dataset analyses were also entered into the Neuro-QOL item library.

Finally, our team drew from additional generic and disease-specific HRQL measures that were identified from in-house sources and a series of literature searches. First, the comprehensive literature review conducted prior to expert interviews and patient/caregiver focus groups was evaluated to determine the extent of research and information regarding quality of life issues for patients with the selected neurological conditions. Relevant generic HRQL questionnaires that were used with patients with the selected conditions were identified and appropriate questions from these instruments were candidates for inclusion in the item library, pending approval of the developers. Similarly, items from targeted HRQL instruments were identified and included in the item library in a similar way.

Table 20. External Neurological Datasets

Name	PI	N	Source	Assessment(s)	Disease (s)	Instrument(s)
Cardiovascular Health Study (CHS)	Olson	5,888	NHLBI	10	Stroke, Cardiovascular Diseases	CES-D (Center for Epidemiologic Studies Depression Scale), Functional status, ISEL (Interpersonal Support Evaluation List), LSI (Life Satisfaction Index), LSNS (Lubben Social Network Scale), Stressful life events
Epilepsy Impact Study	Michael Reed	775 epilepsy, 395 asthma, 362 no chronic health condition	Vedanta Research	1	Epilepsy, Asthma, People without chronic conditions	CES-D, SF-36, QOLIE-89 (Quality of Life in Epilepsy Questionnaire), Sheehan Disability Scale, Adverse Events Profile, Social Concerns Index
European ALS Health Profile Study	Crispin Jenkinson	1300	University of Oxford	Baseline and several follow up time points	ALS	SF-36, ALS Functional Rating Scale
PDQ-39	Crispin Jenkinson	839	University of Oxford	Baseline and several follow up time points	Parkinson's Disease	PDQ-39
Secondary Prevention of Small Subcortical Stroke (SPS3)	Oscar Benevente	550	University of Alabama Computing Center	Baseline and annual follow-ups. Average 4 years of follow-up.	Lacunar stroke	SSQOL (Stroke-specific QOL Scale), Subjective Symptom Assessment Profile for Hypertension
Sonya Slifka Longitudinal MS Study	Sarah Minden	2,156	Abt Associates and NMSS Headquarters in NY	Baseline, 12 months, 24 months, 36 months, 48 months.	Multiple Sclerosis, NMSS members and others, all diagnosed by a physician with MS.	SF-12, MSQLI, ADL and IADL derived from the Medical Expenditure Panel Survey and the National Health Interview Survey
Lomalinda Parkinson's Dataset	Grenith Zimmerman	97	LLUAHSC Department of Surgery	2	Parkinson's Disease	Unified Parkinson's Disease Rating Scale, FACT-G, PDQ-39

3.9.2 Organization and Classification of Items

3.9.2.1 Development of an Item Library. Instruments and items that were identified by Neuro-QOL investigators and expert consultants by literature searches, previous item banking projects and large dataset analyses were delivered to a centralized Neuro-QOL Item Library. Over 3,482 items were entered into the Neuro-QOL Item Library according their elements such as item order, context, time frame, item stem and response options. Please see the example in Table 21.

Table 21. Example of Neuro-QOL Item Library

Item ID	Questionnaire	Item Order	Item Content	Item Stem	Response Options
70	Leeds Multiple Sclerosis Quality of Life (LMSQoL) Scale	8	Tick box according to your perception of the item in question. The time reference used was the preceding month	I have spent evening out with my partner	4=very much; 3=quite a lot; 2=a little; 1=not at all
71	Leeds Multiple Sclerosis Quality of Life (LMSQoL) Scale	9	Tick box according to your perception of the item in question. The time reference used was the preceding month	I have looked after house (cleaning, repair)	4=very much; 3=quite a lot; 2=a little; 1=not at all
72	Leeds Multiple Sclerosis Quality of Life (LMSQoL) Scale	10	Tick box according to your perception of the item in question. The time reference used was the preceding month	My health causes problems with my job	4=very much; 3=quite a lot; 2=a little; 1=not at all
73	Leeds Multiple Sclerosis Quality of Life (LMSQoL) Scale	11	Tick box according to your perception of the item in question. The time reference used was the preceding month	I feel secure in my present job	4=very much; 3=quite a lot; 2=a little; 1=not at all
74	Leeds Multiple Sclerosis Quality of Life (LMSQoL) Scale	12	Tick box according to your perception of the item in question. The time reference used was the preceding month	I feel lonely	4=very much; 3=quite a lot; 2=a little; 1=not at all

3.9.2.2 Binning. Once the Neuro-QOL item library was populated, items were assigned to respective Neuro-QOL domains through an iterative, multi-step process involving at least three domain experts. Two independent raters worked collaboratively to bin items to primary domains followed by agreement and reconciliation by a 3rd reviewer to ensure consistency across domains. Each item was assigned to a primary bin and if applicable (“1”), it was also assigned to a secondary bin (“2”). As the number of items (many redundant) that existed in the

library was large, all items were reviewed to determine if they should proceed through detailed item review/revision/testing and were grouped together according to each domain's hierarchy of sub-domains, factors, and facets. Please see Table 22 for an example of the item binning process.

Table 22. Example if Item Binning Process

Questionnaire	Item Order	Item Content	Item Stem	Emotional Distress	Cognitive Function	Physical Function	Social Function
Quality of Life in Epilepsy for Adolescents (QOLIE-AD-48)	34	The following questions ask about how your epilepsy or medications (antiepileptic drugs) have affected your life in the past 4 weeks. In the past 4 weeks, how often did you …	Feel that epilepsy or medications limited you social life or dating	-- -	---	---	1
Quality of Life in Epilepsy for Adolescents (QOLIE-AD-48)	35	The following questions ask about how your epilepsy or medications (antiepileptic drugs) have affected your life in the past 4 weeks. In the past 4 weeks, how often did you …	Feel that epilepsy or medications limited your participation in sports or physical activities	-- -	---	---	1
Quality of Life in Epilepsy for Adolescents (QOLIE-AD-48)	36	The following questions ask about possible side effects from antiepileptic drugs. In the past 4 weeks, how did you feel …	About how you looked (side effects such as weight gain, acne/pimples, hair change, etc.)	1	---	---	---

3.9.2.3 Winnowing. Once all items were assigned to respective domains, content experts "winnowed" (or systematically removed items) from items pools" because of semantic redundancy, availability of a superior alternative, inconsistency with domain definition, wrong domain assignment, vague or confusing language, cultural/translation relevance, gender inappropriateness, too narrow, or too disease specific. Items selected from winnowing underwent a more thorough review done collaboratively by two domain co-chairs and several outside content experts. Most items needed revision for general consistency across banks. Re-writing or generating new items was done to assure comprehensiveness in measuring the domain; clear, understandable and precise language to experts and respondents; amenable to linguistic translation; and adapted to the data collection and analysis strategies planned.

***3.9.2.4 Qualitative Item Review*.** The qualitative item review (QIR) process began once classification of items ("binning") and selection of items for further review ("winnowing") for a given domain's potential item bank was complete. QIR consisted of two efforts: (1) expert item review (EIR) and (2) patient cognitive interviewing. Although quantitative information was also used to

help determine final item pools for testing, these three efforts represented qualitative approaches to improving and adapting items for administration in a computer-based testing (CBT), computer adaptive testing (CAT) and item-response theory (IRT) framework for Neuro-QOL.

Prior to QIR, Neuro-QOL activities included a comprehensive review of the literature and existing instruments, establishment of a conceptual framework, expert interviews, patient focus groups and the creation of item pools that at that point contained many more potential items than would be possible to administer, in forms not yet ready for administration. Binning and winnowing of the items prior to QIR have ensured relevance to the domain frameworks, and organized and streamlined the item banks by categorizing and paring down the massive quantities of items in each bank. QIR prepares the items for administration by further categorizing, unifying, and re-writing them to produce a set of items that are relevant, optimized, and adapted to the technologies we plan to use (CBT, CAT and IRT). Similar to scale development processes, item preparation through QIR creates new items and adapts existing items based on two key sources: expert opinion (expert item review; EIR) and patients/potential research participants (cognitive interviews). Our previous expert interviews and patient focus groups helped provide input to conceptual gaps in the domain definitions, which led to the identification of new items, especially where it was judged that existing items did not provide adequate coverage. Cognitive interviews in English and Spanish helped ensure that items selected for testing would be understood as intended by respondents, especially those with neurological disorders and/or low literacy. See Table 23 for QIR Tasks.

Table 23. Qualitative Item Review Tasks

Task	Assigned Member(s)
Review and revision of items	Domain co-chairs working collaboratively
Review of revised items	Outside expert consultants
Reach consensus on all items' revisions	Domain co-chairs & expert consultants; translatability assessment
Evaluate conceptual gaps in domain and domain framework	Domain co-chairs & expert consultants
Cognitive assessment for each item	Individual patient cognitive interviews
Construction of new items and & revision of old items based on cognitive interviews	Domain workgroup members
Achieve consensus on response categories	Domain co-chairs & expert consultants
Review of finalized items	Entire domain workgroup
Submission of items to Executive Committee in test-ready form	Domain co-chairs

At least three people reviewed each item for expert item review. Two domain co-chairs worked together to produce a set of preliminary items. Additional outside expert consultants then participated in the item review process. The entire domain workgroup was involved in the final phases of reaching consensus on all items. The final set was submitted to the remaining Executive Committee members who were not actively involved in these efforts.

A proposed model was established for the two domain chairs to meet or speak to categorize and suggest up to 50 candidate items, aiming to reaching consensus on that portion. Some found it useful to work separately, and then reconvene to reach consensus. At the end of these meetings a set of items was selected, reviewed and modified (if necessary), and the 2 domain members submitted the items to additional expert collaborators for review. Expert collaborators (a) signed off on items that appeared to need no further revision, indicating consensus, and (b) suggested revisions to items that still needed improvement. Expert collaborators worked with the domain group members to reach consensus on the final re-written form of each item. The final list of items was sent to translation experts who provided feedback about the translatability and cultural relevance of the proposed list of items, and worked with the domain group to revise items as needed to improve translatability and cultural relevance. Items and conceptual gaps were also concurrently evaluated through patient cognitive interviews. Final item pools were reviewed by adult and pediatric patients with Neuro-QOL conditions, as well as several pediatric caregivers (n=63) during telephone-based cognitive interviews in English and Spanish to assess the content validity of items, clarify concepts, refine language and response options. During interviews patients reviewed each item in a one-on-one semi-structured interview focused on item comprehension and relevance. Patients also identified areas (gaps in domain) for new item development and creation, to which new items were written or revised. This input was used by the domain workgroup to construct new items and revise old items. Domain group members reviewed items reviewed a list of proposed response categories. These suggestions were reviewed by the entire group and modified. Translation experts were consulted to provide feedback about the translatability of each response set. Domain group members achieved consensus on response categories and the entire Domain Workgroup reviewed the final set of items. Domain Co-Chairs submitted the domain items in test-ready form to the Executive Committee after giving final approval.

3.9.2.5 Criteria for Evaluating Items. Because many of the items received by the reviewers had already undergone binning and winnowing (from PROMIS) and were already part of a pre-existing, calibrated item bank (e.g., AM-PAC from Boston University), we didn't anticipate excluding many items. Rather, Neuro-QOL assumed that depending on neurology-specific issues, some items needed at least some level of re-writing, ranging from minor modifications to a complete overhaul.

Items were revised or re-written with the goal of optimization. The principle was to preserve as much as possible of the original item, but to help the item fit within the Neuro-QOL framework for administration. For confusing items, this process offers the opportunity for making them clearer. Neuro-QOL used the following guidelines for re-writing.

Clarity: Items were revised if they were deemed unclear including the clarity of instructions to the respondent, and the clarity of the item text, including singularity of concept. Items were also revised if an item seemed too long, written at a high literacy level, or was written with poor grammar. If an item

appeared too vague to elicit a concrete response, an alternative was suggested. Finally we revised any aspects of the item context, stem, or response options that would have presented significant challenges to translation and cross-cultural applicability. For example, the concept of walking a “block” may not be applicable to non-North American cultures.

Precision: When an item appeared to measure more than one concept it was re-written to break down into one-item-per-concept. Also, when an item was ambiguous and could be interpreted in more than one way, it was re-written so that the ambiguity was resolved.

Acceptability to respondents: We revised aspects of the item context, stem, or response options that seemed to impede one’s ability or willingness as a respondent to provide an informative answer. Examples included cultural or gender biased items, use of colloquialisms or jargon, and potentially demeaning questions (e.g., the stigma item: ‘I am disgusting’ wasn’t received well by many patients during cognitive interviews)

Adaptation to data collection format: CBT and CAT

Format of items: Items were revised to meet the Neuro-QOL format for item stems and contexts. The item stem is usually a standalone statement or question that captures the essence of what is to be measured. For example, "I have to limit my social activity because of my health"; "Has your health interfered with your social activities?" The item context is generally some instructional material, often including the timeframe. For example, "Please indicate how true each statement has been for you during the past 7 days.” Aspects of item context (e.g., time frame, general directions, etc.) and stem were set aside for future use according to Neuro-QOL conventions that will be developed, e.g., displaying it only once in an introductory screen, displaying it on every screen, etc.

Set of preferred response options: Items were revised to fit a Neuro-QOL adopted response set. Most if not all items utilize an accepted response set. Response options were adapted by domain workgroups and it was decided that no bank should have more than 2-3 different response options.

Timeframe: We adopted a 7-day recall period for most item banks (similar to the NIH PROMIS system) because it is typically at the upper limit of ecological validity for recall of specific events (especially for subjective symptoms), yet long enough to allow for enough events to occur. However, for certain banks, we accepted alternative time frames, such as “Lately” for stigma and positive affect and wellbeing or no time frame at all for some of the physical banks.

***3.9.3 Spanish Translation*.** The Neuro-QOL bank items were translated using the FACIT translation methodology [2] as described below:

1) Two simultaneous forward translations: Source items in English are translated into Spanish by two independent professional translators who are native Spanish-speakers from different countries of origin.
2) Reconciled single Spanish translation: A third independent translator, also a native Spanish speaking professional, reconciles the two forward translations by choosing the best of the two forward translations and resolving discrepancies between them.

3) Back-translation: This reconciled version is then back-translated by a native English speaking translator who is fluent in Spanish. The back-translator is blind to the original source English version.
4) Back-translation review: CORE Translations staff compares source and back-translated English versions to identify discrepancies in the back-translations and to provide clarification to the reviewers on the intent behind the items.
5) Expert reviews (3): Three bilingual experts examine all of the preceding steps and select the most appropriate translation for each item or provide alternate translations if the previous translations are not acceptable.
6) Pre-finalization comments: CORE Translations staff assesses the reviewers' comments to identify potential problems in their recommended translations and formulates questions and comments to guide the language coordinator.
7) Finalization: A language coordinator, who worked on the translation development either as a reviewer or as the reconciler, reviews all the information and determines the final translation. The language coordinator also provides a literal back translation and a polished back translation of the final Spanish translation.
8) Formatting, typesetting and proofreading of final questionnaire or item forms by two proofreaders working independently, and reconciliation of the proofreading comments.
9) Pre-testing: The target language version is pre-tested with subjects who are native speakers of Spanish. Each item is debriefed by at least 5 native Spanish-speaking patients in a cognitive debriefing interview to ensure that the meaning of the item is equivalent after translation.

Prior to beginning the translation process, the items were incorporated into a document called an Item History in which each item and its subsequent translations were listed on a separate page. This format makes it possible to focus on the translation item by item, and provides a convenient format for the translators and reviewers to visually compare the different translations and back-translation and to provide comments on the translation of each item. Because of the item history format, the finalized items were subsequently formatted into the layout appropriate to the project for the pre-testing phase.

The first translations occurred in November 2006 of blocks 6, 7, and 8 of adult banks (126 items) and in January 2007 of block 3 of pediatric (60 items). These items were translated for use in cognitive debriefing interviews with Spanish speaking patients during NEURO-QOL item development. In this first round, Kramer Translations conducted all the translation steps using the methodology described previously, and CORE Translations staff reviewed the back-translations and pre-testing finals. Kramer did formatting and proofreading as well. As a result of debriefing with English speaking patients, changes were made to the English items in NEURO-QOL after they had already been translated into Spanish. Some changes were global, such as going from present tense in the first version to past tense in the final version. The NEURO-QOL English items were finalized in September 2007. Before translation, CORE Translations staff conducted a comparison process to see how many of the previously translated items were retained in the banks. This process included identifying items that were modified from the previous version and redundancies between the adult and pediatric

banks. Any discrepancies were resolved in collaboration with the domain chairs, and the items were updated so that the adult and pediatric versions would be consistent.

Once the issues were resolved, the preparation for translation began. The banks related to each domain were placed together to translate as a group. The goal was to have 100 to 150 items in each round of translation to make it manageable for the translators and CORE Translations staff to review. The targeted banks and scales were put together in one item history and translated as a group. It was also decided to include the pediatric banks for each domain in the item history so that everything in each domain could be translated together for purposes of consistency. Item definitions were developed for each item to assist in the translation process. The items previously translated in 2006 were placed in a separate item history and were translated using a minimal methodology to modify the existing translations. The new and old English were included in the item history as well as the translation of the old English item for the translators' reference. The minimal methodology consisted of the following steps: one forward translation modifying the existing translation per revised English, one back-translation, one independent review, and finalization by the language coordinator. CORE Translations conducted these steps with its Spanish translation team. The new items were sent to Kramer Translations to follow FACIT translation methodology. They performed all steps up to formatting.

Another set of items were provided by Boston University from its AM-PAC (Activity Measure for Post Acute Care) instrument, for applied cognition and physical function. A quality review of all new translations was performed by CORE Translations staff. Any translation issues and inconsistencies identified during the quality review process, as well as discrepancies found by the proofreaders, were resolved by CORE Translations and its Spanish translation team. After all translations were completed in the item histories, they were copied and pasted into the Excel file formats provided by the NEURO-QOL team. In order to store the translations and proofread them, both the English formats and the translations were uploaded into the World Server translation memory. After formatting, the translated banks were sent to proofreading, with one proofreader from Kramer translations and the second proofreader from the CORE Translations Spanish team. A number of issues came up during proofreading that had to be addressed. They were as follows: 1). Verb tense inconsistency between the preterit and imperfect past in Spanish when simple past was used in English. The use of simple past or preterit was recommended for all items; 2) Recall period "lately" was inconsistent with the past tense in the items. CORE Translations recommended dropping the recall period for both stigma banks and positive psychological function, but the NEURO-QOL team did not approve the change. As a solution, "lately" was translated as "recently" to allow for the grammatical agreement between the recall period and the past tense used in the items; 3) Difference in translation between BU's and CORE's translations for the AM-PAC phrase "How much difficulty do you currently have…". BU agreed to CORE's revision; 4) Discrepancies between the translation of a few AM-PAC items and the English source. BU agreed to CORE's revisions; 5) Universality considerations: in cases where the Spanish word used for a concept was too country-specific or would not be understood in a specific country, another word was added in parenthesis in an attempt to make it universally understood by all Spanish speakers. This process also affected a few AM-PAC items.

After the proofreading issues were resolved, the items were cognitively debriefed with 30 adults and 30 children. For the adult group, 15 subjects were recruited from the general population and 15 from a clinical population, (10 with any neurological disease and 5 diagnosed with stroke). The banks were separated according to these patient groups:

Clinical: 1. Assisted Device, Personality & Behavior items (Stroke); 2. Fatigue & Weakness, Stigma; 3: Sleep, Stigma.

General: 1. Positive Physical Function, Cognition; 2. Cognition; 3. Upper Extremity, Mobility.

For the pediatric group, 20 subjects were recruited from the general population, and 10 from a clinical population (5 patients diagnosed with epilepsy and 5 with muscular dystrophy), with their items also grouped accordingly:

Clinical: 1. Pain, Wheelchair, Walking Aid, Stigma (MD); 2. Cognition, Fatigue, Stigma (Epilepsy).

General: 1. Emotional; 2. Social; 3. Mobility; 4. Upper Extremity. As a result of IRB approval delays at Children's Memorial Hospital, and difficulty finding subjects with the specific diagnosis, the items assigned to the epilepsy and MD groups were tested with general population. Three interviews had already being conducted with MD patients in California, and the data was combined.

Each subject was asked to first answer the items independently. Completion of the questionnaire was followed by the cognitive debriefing interview, where a Spanish speaking interviewer asked the subject about the meaning of specific words in the items, about the overall meaning of the item, or why they had chosen a specific answer. For some items, the subjects were also asked to consider alternative wording for those items.

All the subjects' comments and suggestions regarding each item were compiled into a Pilot Testing Report (PTR) document, and analyzed by CORE Translations staff to determine if the items were well understood by the Spanish-speaking population. After reviewing their comments and consulting with the Spanish language coordinator, some revisions were made to the translation. CORE Translations staff once again proofread all the banks to ensure that post-testing revisions were made consistently within the same banks as well as across banks. There are cases where the translation varies depending on the target population (pediatric vs. adult): if the pediatric population did not understand a certain word, while the adult population did, the revision was made only to the pediatric items.

Finally, we note that the Depression, Anxiety, Social Role Performance, and Social Role Satisfaction items were translated and tested under the NIH PROMIS study. This was made possible due to their origin within that project, enabling efficient cost-sharing.

3.9.4 Cognitive Interviewing with Patients and Caregivers. After identifying approximately the 50 best items per generic item bank or disease-specific scale, English (n=42) and Spanish (n=21) cognitive interviews were conducted by telephone with 49 adult and 14 pediatric patients with Neuro-QOL conditions, as well as several pediatric caregivers. During these interviews, patients reviewed each item in a one-on-one semi-structured interview that focused on item comprehension and relevance. The interviewer asked questions to assess the content validity of items, concept clarity, language refinement and ease of using the response options. Respondents also identified areas for new item development and creation. When these were "gaps" in the newly created banks and scales, the Neuro-QOL domain experts either identified a relevant item on an existing HRQL questionnaire or within our other item banking projects OR a new item was written to cover the gap.

3.9.5 Field Testing Ready Item Pools for Adults and Children**.** The proposed set of item pools in Table 24 below represents the total number of items per bank or scale for field testing.

Table 24. Field Testing Ready Item Pools

Adult Banks	**Items Per Bank/Scale**
Social Role Performance	49
Social Role Satisfaction	51
Physical Function Screening Item	1
Mobility/Ambulation	37
Assistive Devices	13
Self-Care/Upper Extremity	44
Depression	31
Positive Psychological Function	27
Fear/Anxiety	28
Cognitive Function	45
Applied Cognitive Function	44
Stigma Bank	26
Personality & Behavioral Change	20
Sleep Disturbance	20
Fatigue/Weakness	20
Pediatric Banks	
Emotional Health	46
Social Function	38
PF Screen	1
Mobility	39
Self-Care/Upper Extremity	41
Assistive Devices	32
Stigma Bank	20
Fatigue	13
Perceived Cognitive Function	20
Pain	10

4 CALIBRATION AND VALIDATION TESTING

4.1 *Overview*. Most of the instruments which make up the generic domains (Physical Health, Emotional Health, Cognitive Health and Social Health) were field tested on samples drawn from the US general population, while targeted domains and instruments (fatigue, emotional and behavioral dyscontrol, and stigma) were field tested in a clinical sample consisting of patients diagnosed with stroke, epilepsy, MS, Parkinson's or ALS. Sleep Disturbance was field tested using both the general population and clinical sample. For Wave 1a, the response data were collected by YouGovPolimetrix (www.polimetrix.com). Their standard respondent pool for an internet-based survey is taken from a predetermined panel of people who typically respond to the company's online surveys. Chosen panelists receive modest compensation (under a $10 value) for their participation. Wave 1b data was collected through Greenfield Online, an alternate online paneling organization, offering a similar service to that of YouGovPolimetrix. Greenfield Online was chosen for Wave 1b because their fees proved more economical for this particular sample, while their recruitment methods were fairly similar.

All participants completed a socio-demographic form consisting of approximately 20 auxiliary items measuring global health perceptions, socio-demographic variables including age, income, number of hospitalizations, disability days, use of prescription medication, height, weight, gender, race/ethnicity, relationship status, educational attainment, and employment status. In addition, subjects responded to a clinical form which included a series of health questions about the presence and degree of limitations as they related to multiple neurological conditions affecting adults including stroke, multiple sclerosis, Parkinson's disease, epilepsy and ALS.

Neuro-QOL data collection occurred in two waves: Wave Ia from January 31, 2008 to March 10, 2008 for clinical samples for domains targeted to certain neurological conditions, and Wave Ib from September 11, 2008 to September 24, 2008 for general population for domains generic across neurological conditions as well as general population. From January 15, 2009 to January 30, 2010 Wave II short form testing in clinical samples was conducted to increase the sample size for some of the instrument calibration analyses and to conduct validation studies. Wave II participants were recruited from Cleveland Clinic Foundation, Dartmouth-Hitchcock Medical Center, NorthShore University HealthSystem, Northwestern University Feinberg School of Medicine, Rehabilitation Institute of Chicago, University of Chicago, University of Puerto Rico, and the University of Texas Health Science Center. Validation results will be discussed in subsequent publications. The sampling plan facilitated obtaining item calibrations for the different domain areas, estimating profile scores for varied subgroups, confirming factor structure, and conducting item and bank analyses. Given the large number of items (>500), we knew that participants could not be asked to respond to the full item pool. It was estimated that participants would respond to four questions per minute, with the maximum number of items administered for each respondent approximately 150. This led to a response time on average of 37 minutes.

4.2 *Wave 1a Online Clinical Sample*. The Neuro-QOL Wave 1a adult clinical sample included 553 respondents (see Tables 25 and 26). Please refer to Table 3 for a full breakdown of all demographic variables from both Wave 1a and 1b samples.

Table 25. Initial clinical sample adult enrollment (Wave 1a)

ADULT BANKS/SCALES	Number of Items per form	Conditions	Total Sample Size (All English Speaking)
Socio-demographic Form	20	• Stroke (n=209) • Epilepsy (n=183) • MS (n= 84) • Parkinson's (n= 59) • ALS (n=18)	553
Clinical Form	82		
Stigma Bank	26		
Emotional and Behavioral Dyscontrol Bank	20		
Sleep Disturbance Bank	20		
Fatigue Bank	20		

4.3 *Wave 1b General Population Sample*. The total Wave 1b adult sample included a total of 3,123 respondents composed of English and Spanish-speakers from the general population with no specific targeted diagnosis (English N=2,113; Spanish N=1,010--see Table 26). Each participant was assigned to complete items included in one of four forms. The sample was used primarily for calibrating item parameters and setting the optimum location for establishing the midpoints of the score range for each calibrated item bank as it related to deriving scores. This would enable comparison of item bank scores to general population benchmark values. As described earlier, in an effort to limit response burden, item banks were divided across a series of test forms which were administered to different samples. For each form the total sample size is listed, as well as the specific number of people who completed the items in each particular item bank. The primary demographic characteristics for each form was similar to the total Wave 1b demographics shown in Table 26.

4.4 *Wave II Clinical Samples*. The total Wave II adult sample included a total of 853 respondents accrued from 8 academic medical centers. The sample was used primarily to conduct validation testing on short form versions of the Neuro-QOL instruments. The data was also utilized to improve the quality of the IRT analyses for the Cognition, Physical Function and Sleep Disturbance instruments. The primary demographic characteristics are shown in Table 26.

Table 26. Wave I and II Sample Demographics

	Wave 1a: Adult Clinical Sample	Wave 1b: Adult General Population (English-speaking)	Wave II Adult Clinical Sample
N	553	2,113	581
Age Average (SD)	56.2 (12.8)	52.67 (15.5)	55.2 (14.3)
Gender			
Male	53%	50%	46%
Female	47%	50%	54%
Race			
White	95%	91%	87%
Black/ African American	3%	5.5%	12%
American Indian/ Alaskan Native	4%	1.5%	2%
Asian	1%	3.3%	2%
Native Hawaiian/ Pacific Islander	6%	1.0%	--
Occupation			
Homemaker	11.5%	12%	8%
Unemployed	8%	8%	9%
Retired	37%	31%	30%
Disability	26.5%	10%	34%
Leave of absence	5%	>1%	1%
Full time employed	25%	31%	21%
Part time employed	10%	12%	10%
Full time student	2%	3%	1%
Marital Status			
Married	60%	52%	62%
Widowed	7%	7%	5%
Living with someone	6.5%	7%	5%
Separated	1%	3%	2%
Never Married	11%	17%	16%
Income			
> $20,000	17%	18%	16%
$20-$49,000	35%	45%	35%
$50-$99,000	30.5%	31%	28%
<$100,000	14.5%	11%	21%
Education			
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%

4.5 *Analysis Plan and Item Calibrations*. The data analysis strategy closely followed to Reeve et al [3] mainly including evaluation of unidimensionality and estimation of item parameters using IRT models. A short synopsis of this plan follows. All of the analyses described in this paper were conducted using the Wave I data, except for the Cognition, Physical Function and Sleep Disturbance scores, which combined for Wave I and Wave II data. In addition to descriptive statistics and item-total correlations, factor analytic approaches were used to evaluate dimensionality of items of each domain. Confirmatory factor analysis (CFA) that accounts for categorical (ordinal) data was often

run without first conducted exploratory factor (EFA) analysis because a model for domain unidimensionality was designed into the item development. When EFA was used, samples were randomly divided into two datasets (one for EFA and the other for subsequent CFA). Since a one-factor model is typically statistically rejectable (chi-square statistic) when a large sample size is used, the focus was on practical fit indices such as the comparative fit index (criterion: CFI >0.90), RMSEA (criterion: < 0.10), factor loadings (criterion: >0.3), and average absolute residual correlations (criterion: <0.15). When a clinically or theoretical meaningful model is available, a bi-factor model was implemented to better describe the domain structure. A bi-factor model allows each item to load on a general factor and one group factor. The (squared) loadings indicate the proportion of variance of the item that is accounted for by the two factors,. Those variance components are independent and measure two different variables using independent variance components. Thus, the group factors are orthogonal to each other and the general factor, and are defined by item content facets. Fitting a bi-factor model to the data allowed for the evaluation of the degree to which using an item set to scale individuals on a common factor is distorted by the presence of small secondary group factors). Samejima's Graded Response Model (GRM) as implemented in MULTILOG was used for IRT related parameter estimations for items that meet the unidimensionality requirements. GRM is a polytomous IRT model which is specifically designed for use with items with ordered categories.

We evaluated parameters stability on gender, education and age. An item displays differential item functioning (DIF) when probabilities of responding in different categories differ by population for the same underlying level of the attribute. Items can be evaluated for DIF by contrasting the IRT difficulty or location (b_i) and slope (a_i) parameters between two groups, which in this context relates to the major demographic groups represented in the Neuro-QOL sample. IRT-based hierarchical ordinal logistic regression (OLR) approach as implemented in LORDIF [4] was used for evaluation of DIF. In this approach a series of logistic models predicting the probability of item response were run and compared. The independent variables in Model 1 are the trait estimate (e.g., raw scale score), group and the interaction between group and trait. Model 2 included main effects of trait and group, and Model 3 included only the trait estimate. Non-uniform DIF was detected if there was a statistically significant difference in the likelihood for Model 1 and Model 2, and uniform DIF is evident if there is a significant difference in the likelihoods for Models 2 and 3.

These results were discussed and decisions were made regarding each item. Typically, a first wave of item “cuts” was made; whereby the most problematic items were eliminated and the reduced-length item pools were subjected to follow up analyses to help arrive at decisions regarding each item. Through this process of iterative analysis and discussion with content (domain) experts, item-by-item level decisions were made determining whether an individual item should be: (1) calibrated and included in the bank, (2) not calibrated but retained for possible future calibration (e.g., items consistent with the domain being measured but having local dependence, responses concentrated in few of the available response options), or (3) excluded from further consideration (e.g. outside of concept; problematic item wording). For complete results of Wave I calibration results, see Gershon et al.[5]

4.6 *Development of 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), we ranked items by the amount of information they provided across the range of what was being measured (e.g., applied cognition; social function; physical function). 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. These item rankings (information and CAT) were then set aside while ten doctoral level clinical and measurement experts (3 neurologists; 4 clinical psychologists; 1 occupational therapist, 1 social worker and 1 neuropsychologist reviewed each of the candidate items for their relevance and appeal based on item content only (item performance was not shared with experts prior to their ratings). Experts were directed to identify the five most-preferred and five least-preferred of the items in the calibrated bank. Individual preferences of each rater were then presented along with item performance statistics. With this information tabulated, we (DC, DV, SC, J-SL, CN, DM, NR) identified items with strong psychometric characteristic (fit to the IRT model; highly informative; selected early by CAT) and high appeal to clinical raters, to the greatest extent possible. We discussed marginal item choices (high clinical appeal but relatively weak psychometric performance; marginal clinical appeal but good psychometric performance) until we reached consensus regarding item inclusion or exclusion from the short. We also considered two other goals: one was respondent burden, so that if one of two nearly-equal items was worded consistently (e.g., same response options) with 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), assuming they were calibrated 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 capability for Neuro-QOL to future cross-walk to the PROMIS item banks.

4.7 *Wave II Clinical Validation Study*. We conducted a second phase of field testing, from January 2009 through June 2010, to evaluate the reliability, validity and responsiveness of Neuro-QOL short forms and scales in clinical neurology populations. A total of 581 adult and 113 pediatric patients were recruited to reflect the five adult and two pediatric neurological conditions targeted by Neuro-QOL. Proxies for stroke (N=84) and the two pediatric samples (N=113) also completed forms. Administration of Neuro-QOL Short Forms and clinical validation measures (both cross-disease and disease-specific), physician ratings and chart review was conducted at baseline and at a 180-day follow up (to assess responsiveness). Test-retest reliability of the Neuro-QOL Short Forms was evaluated at 7 days. Table 27 lists the number of patients with each respective neurological condition (and proxies) who completed each assessment.

Table 27 – Field Testing/Clinical Validation Sample

	Number completing assessment		
	Baseline	7-day	180-day
Multiple Sclerosis	161	125	132
Parkinson's disease	120	116	108
Adult Epilepsy	119	119	109
Stroke	101	95	90
Stroke Proxies	84	78	73
ALS	80	77	59
Pediatric Epilepsy	62	60	56
Pediatric Epilepsy Proxies	62	60	56
Muscular Dystrophy	51	48	48
Muscular Dystrophy Proxies	51	48	48
Total:	891	826	779

4.7.1 Methods

***4.7.1.1 Participating Sites*.** Participants were recruited from several clinical sites, including: Children's Memorial Hospital of Chicago, Cleveland Clinic Foundation, Dartmouth-Hitchcock Medical Center, NorthShore University HealthSystem, Northwestern University Feinberg School of Medicine, Rehabilitation Institute of Chicago, University of California – Davis, University of Chicago, University of Puerto Rico, and University of Texas Health Science Center.

***4.7.1.2 Site Procedures*.** Each accrual site had a coordinator who assumed overall responsibility for the project at that particular site. All procedures were approved by the NorthShore University HealthSystem Institutional Review Board (IRB) as well as IRBs at each respective institution. Site coordinators identified, enrolled and conducted assessments with eligible participants according to criteria and procedures specified in the Manual of Procedures. Because our goal was to produce a generalizable measurement platform, eligibility criteria were broad. Table 28 lists our general inclusion/exclusion criteria.

4.7.1.3 Inclusion/Exclusion Criteria.

Table 28. Wave II General Inclusion/Exclusion Criteria

INCLUSION CRITERIA						EXCLUSION CRITERIA
Group	**Age**	**Gender**	**Language**	**Diagnosed Neurological Condition**	**Proxy**	
Children	Epilepsy: 10-18 MD: 10-21*	Proportional breakdown of males and females according to incidence rates of respective conditions	English	Epilepsy, Muscular Dystrophy	Proxies (primary care givers**) of children with epilepsy or muscular dystrophy	• Younger/older than the age limits • Non-English speaking • Cognitive impairment such that it would prevent informed consent and/or completion of test items with the assistance of an interviewer (as determined by recruiting staff). • Does not have a proxy (for adults with stroke or children with epilepsy or muscular dystrophy)
Adults	>18	Proportional breakdown of males and females according to US incidence rates	English	Stroke, MS, ALS, Parkinson's Disease, Epilepsy	Proxies of patients with stroke	

*Due to the nature and developmental impact of muscular dystrophy, participants may be ≤21 years of age to meet eligibility requirements.
** A spouse, parent, adult child, or significant other living with the participant was identified as a primary caregiver.

Each disease condition also had special considerations regarding enrollment and inclusion/exclusion criteria. For example, the stroke team considered whether to explicitly include language dysfunction as a requirement for a small sub-set, but a final decision was that this will be pervasive enough within the larger sample to not warrant a separate block of participants. The Barthel Index was used by stoke and MS teams to monitor enrollment to make sure there was adequate representation at different severity levels (e.g., severe – at least 10-20% in each block). Parkinson's disease used Hoehn and Yahr staging to monitor enrollment to make sure there was adequate representation at different severity levels. The epilepsy team included patients with severe (e.g., 2 seizures per month) to mild (no seizure within the past year) seizures, however non-epileptic seizures were excluded. The MS team also included a percentage of patients who had an exacerbation in the past month. The ALS team used the ALSFRS to monitor enrollment to make sure there was adequate representation at different severity levels and tried to recruit a certain number of patients in their 20s, 30s and 40s as well as with devices. The MD team included patients spanning ambulatory, non-ambulatory and on ventilation. Due to the nature and developmental impact of

muscular dystrophy, participants could be ≤21 years of age to meet eligibility requirements.

***4.7.1.4 Recruitment and Testing*.** Various recruitment methods were utilized including: 1) approaching patients in clinics and 2) mailing letters of invitation to physician-identified patients informing them that someone would contact them about the study at their next clinic appointment. Informed consent or assent (for pediatric participants) was obtained from each subject and covered all three assessments (baseline, 7 days, and 180 days). There was a 5-9 day window for the test-retest assessment and a 5-7 month window for the responsiveness assessment. After a patient was identified and approached, the site coordinator arranged a meeting to introduce and describe the study, confirm eligibility, explain participants' rights, and obtain informed consent and HIPPA authorization if the eligible participant was interested. Site personnel then either administered the baseline evaluation at that time or else scheduled it for another time. Baseline evaluations, consisting of Neuro-QOL instruments, concurrent validity measures, and sociodemographic and clinical data forms, lasted approximately 90 minutes. Some measures, including the Neuro-QOL instruments, were administered by Computer Assisted Self Interview. Other measures were administered by study staff (e.g., performance-based cognitive measures). Medical professional ratings and chart review were also conducted at baseline and as part of the 180-day follow up. Participants were reimbursed according to local IRB-approved standards.

4.7.1.5 Overview of Measures and Administration Schedule. Demographic and clinical data were collected in addition to several cross disease and disease specific measures. See Table 29 below.

4.7.1.6 Cross Disease Measures

General Forms

Demographic and clinical data were collected with two forms:

Socio-demographic form. This form provides patient characteristics (e.g., age, gender, race, ethnicity and education). This information was collected at baseline via chart review and/or face-to-face interview.

Clinical information form. This form records disease specific information (e.g., date of diagnosis, treatments) for each participant. It was gathered via chart review and through interviews with patients and/or parents at baseline and 180-day follow-up interviews.

Neuro-QOL Short Forms. All short forms provided raw scores which were converted to T-Scores; with a T = 50 indicating average function compared to the reference population and a standard deviation of 10. Neuro-QOL T-scores referenced to a general population sample are indicated by GPT (General Population T-Score) while those referenced to a clinical sample are indicated by CT (Clinical T-Score).

Table 29. Cross Disease and Disease Specific Measures

Data and Outcomes	# items	Time required (minutes)	Baseline	7 Days	6 month	Mode of Administration
ADMINISTERED ACROSS CONDITIONS						
Socio-demographic Form	9	<5	X	--	--	Self-report
Clinical Information Form	19	<5	X	--	X	Interviewer
Barthel Index*	10	<5	X	--	X	Interviewer
Instrumental Activities of Daily Living*	8	<5	X	--	X	Interviewer
Karnofsky Performance Status	1	<2	X	--	X	Medical Prof Rated
Oral Digit Symbol Modalities	0-133	<3	X	--	X	Self Report / Interviewer
Digit Symbol Coding	0-133	<3	X	--	X	Self Report / Interviewer
Symbol Search	0-60	<3	X	--	X	Self Report / Interviewer
Global HRQL Question	1	<2	X	--	X	Self-report
Pain Question	1	<2	X	--	X	Self-report
Global Rating of Change Scores	1	<2	--	--	X	Self-report
EQ-5D	15	<3	x	--	x	Self-report
PROMIS Global Health Scale	10	<2	x	--	x	Self-report
Neuro-QOL Short Forms and Scales	±100	45-60	X	X	X	Self-Report
ADMINISTERED IN STROKE						
American Heart Association Stroke Outcomes Classification	3	<3	X	--	x	Medical Prof Rated
Stroke Specific-QOL Scale	49	<10	X	--	X	Self-report
ADMINISTERED IN PARKINSON'S DISEASE						
Montreal Cognitive Assessment	11	<15	x	--	x	Interviewer
Hoehn and Yahr Staging	1	<2	X	--	X	Medical Prof Rated
UPDRS	42	10	X	--	X	Self-report / Interviewer
PDQ-39	39	<10	X	--	X	Self-report
PHQ-9	9	<3	x	--	x	Self-report
ADMINISTERED IN ADULT EPILEPSY						
LSSS	12	6	X	--	X	Self-report / Interviewer
QOLIE-31	31	10	X	--	X	Self-report
LAEP	19	5	X	--	X	Self-report
ADMINISTERED IN MULTIPLE SCLEROSIS						
MSFC (without PASAT)	4	<5	X	--	X	Self-report / Interviewer
FAMS	44	10	X	--	X	Self-report
ADMINISTERED IN AMYOTROPHIC LATERAL SCLEROSIS						
ALSFRS-R	13		X	--	X	Medical Prof Rated
ALSAQ	40	10	X	--	X	Self-report
ADMINISTERED IN PEDIATRIC EPILEPSY & MUSCULAR DYSTROPHY						
PedsQL + MFS	31 (23+ 18)	15	X	--	X	Self-report

*Not administered in pediatric conditions

General Function – Adults Only

Barthel Index. The Barthel Index was developed by Mahoney and Barthel [6] and is one of the best known and most widely used instruments to assess basic activities of daily living (ADL). The Barthel Index assesses the degree of independence a patient has in performing various self-care and mobility ADL tasks. The weighted ordinal scale assesses 10 items of ADL in the following subgroups: personal care (including eating), dressing, personal hygiene and bathing, continence of urine and stool, mobility (including transfer from a bed and toilet), walking, and steps. The index has high test-retest reliability (r=0.89), inter-rater reliability (r>0.95),[7] and internal consistency (Cronbach's alpha = 0.98).[8] We administered this by standardized interview.

***Instrumental Activities of Daily Living Scale*.** The Lawton Instrumental Activities of Daily Living Scale,[9] is an interviewer administered measure which includes 8 items: telephoning, shopping, food preparation, housekeeping, laundry, transportation, medications, and handling finances. Each task is graduated in a 3- or 4-level scale. The scale measures performance in contrast to ability.

General Function – Adults and Children

Karnofsky Performance Status Scale (KPSS).[10] The KPSS is a rating of functional impairment and offers a simple if coarse breakdown of activity level across patients regardless of diagnosis. KPSS criteria are based on descriptive categories from 0-100. Ratings were made by providers.

Cognitive Function – Adults and Children

Oral Digit Symbol Modalities. [11] This is a test of speed of information processing, but is also thought to assess visual acuity and figural memory. A timed coding task using a key as reference, examinees pair specific numbers (0-9) with designated geometric figures that are matched up in the key; examinees attempt to complete as many matches as quickly as possible in 90 seconds. Written and oral forms are highly correlated (in normal adults >.78). Because some participants may have greater motor deficits compared to others, we administered the oral version.

Symbol Search. [12] A test of mental speed, this is a timed orthographic measure of visual attention, scanning, and motor speed. Participants must determine if a target nonsense figure is present in a string of figures and mark a corresponding "yes" or "no" box presented at the end of each item.

Digit Symbol Coding. [12] This is a timed paper/pencil symbol substitution task of mental, visual and motor speed. Using a key of paired numbers and symbols, participants must draw corresponding nonsense symbols below rows of numbers.

Health Related Quality of Life – Adults (including proxies) and Children

EQ-5D. [13;14] This is a 15-item self-report measure of health status developed by the EuroQoL Group in order to provide a simple, generic measure of HRQL for clinical and economic appraisal. Applicable to a wide range of health conditions and treatments,

it provides a simple descriptive profile and a single index value for health status. Domains include: mobility, self-care, usual activities, pain/discomfort and anxiety/depression.

PROMIS Global Health Scale. [15] Global health refers to evaluations of health in general rather than specific elements of health. The PROMIS global health items include global ratings of the five primary PROMIS domains (physical function, fatigue, pain, emotional distress, social health) and general health perceptions that cut across domains. It can be scored into a Global Physical Health component and Global Mental Health component. Global items allow respondents to weigh together different aspects of health to arrive at a 'bottom-line" indicator of their health status. Global health items have been found to be consistently predictive of important future events such as health care utilization and mortality.

Global HRQL Question. [16]A single item from the Functional Assessment of Chronic Illness Therapy (FACIT), "I am content with the quality of my life right now," was used as a global measure of quality of life.

Health Related Quality of Life – Children and Pediatric proxies

Pediatric Quality of Life Inventory, Multidimensional Fatigue Scale (PedsQL™-MFS)[17;18] The PedsQL - MFS is a self-report measure consisting of both a general quality of life measure (PedsQL™) and a fatigue specific measure (MFS). The PedsQL™ is designed to measure core health dimensions in children from 2 to 18 years old. The measure consists of 23 items in four scales: physical functioning, emotional functioning, social functioning, and school functioning. Children/Teens completed a self-report assessment. Proxies completed the parent/caregiver form. The MFS consists of 18 items across three domains: general fatigue (6 items), sleep/rest fatigue (6 items), and cognitive fatigue (6 items).

Pain – Adults (including proxies) and Children

Pain question. A single (0-10) item that asks patients to rate, from "none" (0) to "the worst pain you can think of ("10"), the severity of their worst pain during the past week.

Responsiveness – Adults and Children

Karnofsky Performance Status Scale (KPSS).[10] Described above.

Global rating of change. This measurement strategy assumes that a patient can judge whether over the course of a specified period, their self-reported health status has changed. Typically, such questions require patients to remember a prior health state and compare it to how they are currently feeling.[19;20] In this study, participants were asked to rate how much their Physical, Emotional, Cognitive, Social/Family and Symptomatic Well-being and their overall quality of life had changed over the past 6 months according to the following scale: +3 = "Very much better" to -3 = "Very much worse". Such global transition ratings have the advantage of being easy to interpret and they enhance the interpretability of HRQL scores when found to be correlated with the

target instrument. For instance, if the correlation between a global rating of change and the change score on a target instrument is over 0.5, the validity of the target instrument is supported. Global transition ratings have been widely used in HRQL outcome assessments to augment the interpretation of HRQL scores.[21-23] Proxies completed a proxy version of this measure.

4.7.1.7 Disease-Specific Measures

STROKE

Stroke Specific Quality of Life (SS-QOL) scale. [24] The SSQOL is a 49 item self-report measure containing domains of energy, family roles, language, mobility, mood, personality, self-care, social roles, thinking, vision, upper extremity function and work-productivity. Items are scored on a 5-point Likert scale. Although relatively new, initial psychometric properties are good.

The American Heart Association Stroke Outcome Classification (AHA.SOC). [25;26] The AHA.SOC score provides a mechanism to comprehensively document stroke impairments and disabilities in a single summary stroke score. The system can be used by healthcare providers to reliably assess recovery, measure responses to treatment, and describe the long-term impact of stroke on survivors.

AMYOTROPHIC LATERAL SCLEROSIS (ALS)

Amyotrophic Lateral Sclerosis Assessment Scale *(ALSAQ*: [27], [28], [29]. The ALSAQ is comprised of 40 items across 5 subscales tapping the major domains affected by ALS. The subscales include physical mobility, activities of daily living, eating and drinking, communication and emotional functioning. All 40 items can also be summed together to obtain a total score for ALS QOL. Recently, the scale authors published data on the score differences that might be considered to meaningfully differentiate between subgroups or within groups of subjects over time.[30] This makes the ALSAQ particularly valuable for evaluating the convergent validity and responsiveness of the Neuro-QOL item banks.

Amyotrophic Lateral Sclerosis Functional Rating Scale-Revised *(ALSFRS-R* [31]. The original scale, the ALSFRS, has 10 items that assess activities of daily living, such as speech, swallowing, handwriting, and dressing and hygiene that are specifically affected by the disease. In 1999, three additional items were added to better assess respiratory function. Both the original and revised versions have been used successfully as clinical trial outcome measures.[32] Because of the importance of respiratory problems in the ALS population, we administered the 12-item ALSFRS-R.

MULTIPLE SCLEROSIS (MS)

Functional Assessment of Multiple Sclerosis (FAMS). The FAMS was developed by Cella and Aarnoson and includes 44 questions, divided into six subscales: mobility, symptoms, emotional well-being (depression), general contentment,

thinking/fatigue, and family/social well-being. Fifteen un-scored questions are included because of their clinical value.

Multiple Sclerosis Functional Composite Measure (MSFC). The MSFC was developed as an outcome measure by the National MS Society's Clinical Outcomes Assessment Task Force to address the poor reliability and sensitivity of available MS rating scales [33]. The MSFC consists of three objective quantitative tests of neurological functioning : arm, leg and cognitive function. Arm function is assessed with the nine-hole peg test; leg function with the timed 25-foot walk, and cognitive function with the Paced Auditory Serial Addition Test (PASAT) (being substituted with Oral Symbol Digit test for this study). The MSFC correlates with MRI parameters,[34-36] measures of disability,[37-39] and has predictive validity.[38;40;41] MSFC scores are sensitive to change[42]. It demonstrates excellent intra-rater (ICC =.97) and inter-rater (ICC =0.95 - 0.96) reliability[38;43] for technicians trained with standardized procedures. Scores on the three MSFC components are transformed into Z scores, and then combined into a total MSFC Z score, providing a continuous scale of measurement.

The MS Performance Scales is a medical professional reported measure of MS-related disability. The Performance Scales measure disability in eight domains of function: mobility, hand function, vision, fatigue, cognition, bladder/bowel, sensory, and spasticity. The construct and criterion validity of the subscales of the Performance Scales has been established. [44]

PARKINSON'S DISEASE

Montreal Cognitive Assessment (MoCA).[45] Designed as a rapid screening instrument for mild cognitive dysfunction, it assesses different cognitive domains: attention and concentration, executive functions, memory, language, visuoconstructional skills, conceptual thinking, calculations, and orientation. Scores range from 0-31, with scores below 26 considered abnormal.

Parkinson's disease Questionnaire-39 (PDQ-39). [46;47] The thirty nine items of this self-report measure assess eight dimensions: mobility, activities of daily living, emotional well-being, bodily discomfort, stigma, social support cognition and communication. Scale and summary scores are available, ranging from 0-100, with higher scores indicating greater problems.

Unified Parkinson's Disease Rating Scale (UPDRS).[48] The UPDRS is the most widely used measure of disability and impairment associated with PD. It is a composite scale consisting of 4 parts: Mentation, Behavior and Mood (UPDRS mental score); ADLs (UPDRS ADL score), Motor Function (motor score); and Complications of therapy. The first 3 subscales are quantitative five point scales (0-4). The complications of therapy is a yes/no scale. For this study, UPDRS Motor Function scoring was modified as follows: only the most affected side or body part was rated. All ratings were made by physicians or other medical personnel.

Hoehn and Yahr staging. [49] The Hoehn and Yahr staging consists of 5 disease severity categories ranging from 0.0 (no signs of disease) to 5.0 (wheelchair bound or bedridden unless aided). The staging was obtained through chart review or through direct contact with the patient's physician or other medical personnel.

Patient Health Questionnaire-9 (PHQ-9). [50] This is a 9-item subset of the PHQ, and assesses self-reported depression. The nine items of the PHQ-9 come directly from the nine DSM-IV signs and symptoms of major depression.

ADULT EPILEPSY

Quality of Life in Epilepsy-31(QOLIE-31).[51;52] The QOLIE-31 is an HRQL survey for adults (>18) with epilepsy. Derived from the QOLIE-89, this scale contains domains that include seizure worry, emotional wellbeing, energy/ fatigue, cognition, medication effects, social effects, health status and overall quality of life. Good psychometric evidence has been reported in previous studies.

Liverpool Seizure Severity Scale (LSSS). The LSSS is a 12 item scale that assesses experiences during and immediately after a seizure such as loss of consciousness and post-ictal confusion. Each item is scored on a Likert scale, with higher scores indicating greater seizure severity. Reported test retest reliabilities range from 0.74 – 0.80.[53;54] A modified scoring system requires patients to rate only their most severe seizure and demonstrates adequate reliability, construct validity and responsiveness to change.[55]

Liverpool Adverse Events Profile (LAEP). [56]The LAEP is a 19 item self-report scale that assesses the frequency of antiepileptic drug side effects. Using a 4-point Likert scale (1= never a Problem – 4=always a problem), scores are summed to create a total score (ranging from 19-76, higher scores indicating more symptoms).

***4.7.1.8 Statistical Analyses*.** The following analyses were conducted for all clinical groups.

1. Means, standard deviations, and other distributional statistics were calculated for all scores at the baseline and follow-up assessments.
2. Internal consistency reliability - Internal consistency analyses were performed for each short form using Cronbach's alpha coefficients.
3. Test-retest reliability - Intraclass correlation coefficients and corresponding 95% confidence intervals were calculated to assess the test-retest reliability of the Neuro-QOL measures using the baseline and 7-day assessments.
4. Concurrent validity was assessed at baseline by Spearman rho correlations between Neuro-QOL short forms and disease-specific and cross-disease measures.
5. Known groups validity was evaluated at baseline by comparing mean Neuro-QOL short form scores between patients grouped by clinical anchors such as disease severity. Analysis of variance (ANOVA) was used to test for differences between groups. Effect sizes (mean difference / pooled standard deviation) were calculated to aid in interpretation of group differences.
6. Responsiveness -To demonstrate the sensitivity of the Neuro-QOL measures for detection of change, we evaluated general linear models using each patient's change score. We conducted responsiveness analyses on the Neuro-QOL banks using several criteria for change. One criterion used across all adult conditions was the Karnofsky Performance Status, and another was the self-reported Global Rating of Change (GRC) described above. Here we report the results from the GRC-based change. Beginning with the 7-level GRC (range: +3= very much better; 0 = about

the same; -3 = very much worse), we collapsed the three "better" categories into one, and the three "worse" categories into one, leaving three categories ("better;" "about the same;" "worse"). These three categories were compared using one-way analysis of variance followed by least significant difference testing of adjacent groups when the overall F statistic was significant. For each analysis, we required that at least 5 patients be represented in each of these three categories. If fewer than five patients were represented in a category, it was collapsed with the adjacent category and the two remaining groups were compared using a t-test. There were six GRC questions. Five of them queried patients specifically about change in Physical well-being, Cognitive Well-Being, Emotional well-being, Social/Family Well-being, and Disease-related Symptoms. The sixth GRC item asked about overall quality of life. The list below indicates which of the 13 adult item bank change scores were compared across GRC categories:

Physical well-being:	Upper Extremity and Lower Extremity Function; Fatigue; Sleep Disturbance
Cognitive well-being:	Applied Cognition (General Concerns and Executive Function)
Emotional well-being:	Depression; Anxiety; Stigma; Positive Affect and Well-Being; Emotional and Behavioral Dyscontrol
Social well-being:	Social Function (Ability to Participate in Social Roles and Activities and Satisfaction with Social Roles and Activities); Stigma
Symptoms:	Fatigue; Sleep Disturbance; Emotional and Behavioral Dyscontrol; Depression; Anxiety
Overall:	ALL

This resulted in 32 planned comparisons for adult clinical validation sample (no adjustment made for multiple comparisons). Results for these responsiveness analyses are presented below. Only those that achieved statistical significance will be summarized.

4.8 Wave II Results

4.8.1 RESULTS - STROKE SAMPLE

***4.8.1.1 Sample Characteristics*.** One hundred and one (101) subjects were recruited from 5 centers. Participants were primarily male (55%), white (73%), and non-Hispanic (90%) with average age=59 years (SD=14). Fifty-seven percent were married, 73% had a high school or greater education. Thirteen percent were retired, 33% on disability and 19% were employed either full or part time. Average time post-stroke was 5.4 years (SD=5), with 22% reporting no or minimal deficits, 58% mild/moderate deficits and 20% severe deficits. The primary stroke type was an infarction (71%).

As shown in Table 29, respondents reported worse cognitive and physical function and social well-being than the general population reference group, but more

positive affect and well-being. When compared to a clinical reference group, they reported less depression, fatigue and sleep disturbance, better emotional and behavior control and average stigma.

***4.8.1.2 Reliability*.** Table 30 shows that the internal consistency and 1 week test-retest reliability of the short forms is high, with Cronbach's alphas ranging from .78 to .95 and ICCs ranging from .73 to .94.

Table 30. Descriptive and reliability statistics for Neuro-QOL short form T-scores

Neuro-QOL Short Form	N_{items}	$N_{persons}$	M_{GPT}	M_{CT}	*SD*	α	T-R ICCs**
Positive Affect & Well Being*	9	100	54.92		8.02	.94	.83
Applied Cognition – General Concerns*	8	100	43.70		8.58	.95	.82
Applied Cognition – Executive Function*	8	101	43.67		10.48	.93	.88
Lower Extremity (Mobility)*	8	89	42.73		7.98	.87	.94
Upper Extremity (Fine Motor, ADL)*	8	101	38.45		9.38	.82	.88
Ability to Participate in Social Roles and Activities*	8	100	46.08		7.09	.93	.87
Satisfaction with Social Roles and Activities*	8	100	45.30		5.49	.83	.73
Depression	8	100	47.23		7.48	.94	.81
Anxiety	8	100	50.82		6.61	.90	.76
Stigma	8	100		52.24	8.52	.91	.82
Fatigue	8	100		45.03	8.78	.93	.83
Sleep Disturbance	8	99		46.33	8.25	.78	.76
Emotional and Behavioral Dyscontrol	8	99		45.58	8.47	.89	.79

*For these banks, a high score indicates better function; for all other banks a high score indicates worse function
**Time 1 (baseline) vs. Time 2 (7 days)
M_{GPT} – Mean General Population T-Score; M_{CT}- Mean Clinical T-Score

4.8.1.3 Validity. Table 31 shows Spearman rho correlations between Neuro-QOL short form T-scores and stroke specific measures. Table 32 presents Spearman rho correlations between Neuro-QOL short form T-Scores and cross-disease measures.

Table 31. Correlations for Neuro-QOL short form T-scores with stroke-specific measures

Neuro-QOL Short Form	AHA SOC Number of Neurological Domains Impaired	AHA SOC Severity of Impairment	AHA SOC Level of Function	SS-QOL Total Score
Positive Affect & Well Being	-.17	-.28**	-.33***	.61***
Applied Cognition – General Concerns	-.19	-.31**	-.17	.62***
Applied Cognition – Executive Function	-.36***	-.34***	-.28**	.51***
Lower Extremity (Mobility)	-.23*	-.48***	-.44***	.69***
Upper Extremity (Fine Motor, ADL)	-.33***	-.60***	.54***	.65***
Ability to Participate in Social Roles and Activities	-.34***	-.40***	-.44***	.72***
Satisfaction with Social Roles and Activities	-.18	-.35***	-.39***	.66***
Depression	.19	.30**	.36***	-.66***
Anxiety	.14	.13	.09	-.53***
Stigma	.28**	.40***	.35***	-.59***
Fatigue	.06	.16	.27**	-.59***
Sleep Disturbance	.09	.17	.17	-.50***
Emotional and Behavioral Dyscontrol	.11	.18	.10	-.54***

*p < .05; **p < .01; ***p < .001

Table 32. Correlations for Neuro-QOL short form T-scores with cross-disease measures

Neuro-QOL Short Form	Barthel Index	Lawton IADL Scale	Symbol Digit Modalities # Correct	Symbol Search Raw Score	Digit Symbol Coding # Correct	PROMIS Global Physical	PROMIS Global Mental	Pain Scale (0-10)	EQ-5D Index Score	Global HRQL (0-4)
Positive Affect & Well Being	.36***	.24*	.28**	.23*	.14	.46***	.66***	-.26**	.38***	.52***
Applied Cognition – General Concerns	.29**	.29**	.16	.14	.12	.18	.41***	-.11	.25*	.26**
Applied Cognition – Executive Function	.25*	.34***	.34***	.31**	.28**	.26*	.46***	-.18	.35***	.28**
Lower Extremity (Mobility)	.66***	.44***	.35***	.38***	.32**	.62***	.33**	-.36***	.62***	.42***
Upper Extremity (Fine Motor, ADL)	.65***	.42***	.34***	.38***	.35***	.47***	.38***	-.16	.59***	.36***
Ability to Participate in Social Roles and Activities	.44***	.43***	.21*	.22*	.17	.56***	.58***	-.30**	.54***	.48***
Satisfaction with Social Roles and Activities	.45***	.31***	.22*	.26*	.21*	.56***	.49***	-.43***	.55***	.49***
Depression	-.39***	-.21*	-.20	-.24*	-.04	-.48***	-.66***	.34***	-.46***	-.49***
Anxiety	-.17	-.15	-.01	-.03	.10	-.39***	-.55***	.31**	-.31**	-.36***
Stigma	-.35***	-.20*	-.18	-.20	-.14	-.31**	-.45***	.24*	-.32***	-.52***
Fatigue	-.43***	-.30**	-.22*	-.26*	-.03	-.63***	-.49***	.36***	-.38***	-.38***
Sleep Disturbance	-.22*	-.12	-.21*	-.22*	-.09	-.39***	-.40	.27**	-.24*	-.34***
Emotional and Behavioral Dyscontrol	-.19	-.05	-.05	-.03	.05	-.25*	-.48***	.22*	-.29**	-.41***

*p < .05; **p < .01; ***p < .001

4.8.1.4 Known Groups Validity. AHA severity level was used to split the sample into 3 groups: no/minimal neurological deficit; mild/moderate neurological deficit; severe neurological deficit. These groups differed significantly on all Neuro-QOL short forms except Anxiety, Fatigue, Sleep Disturbance and Emotional and Behavioral Dyscontrol. Effect sizes ranged from -.68 to 2.55.

4.8.1.5 Responsiveness. Of the 32 planned comparisons, 15 were statistically significant and one exhibited a trend toward significance, in the predicted direction.

Physical Well-Being: Of the four planned comparisons [Lower Extremity Function-Mobility, Upper Extremity Function - Fine Motor, ADL, Fatigue, and Sleep Disturbance] three were statistically significant, all in the predicted direction. Specifically, significant differences were observed in Lower Extremity Function – Mobility between patients who reported worsening at six months with those who reported improving in this domain ($F=6.11$, $p<.01$). Similarly, significant differences were observed in Upper Extremity Function - Fine Motor, ADL ($F=6.83$, $p<.01$) and Sleep Disturbance ($F=4.08$, $p<.05$) between patients who reported worsening at six months and those who reported staying the same or improving in this domain.

Social/Family Well-Being: Of the three planned comparisons [Ability to Participate in Social Roles and Activities, Satisfaction with Social Roles and Activities, Stigma] all three were statistically significant in the predicted direction. Specifically, significant differences were observed in Ability to Participate in Social Roles and Activities ($F=3.76$, $p<.05$) and Stigma ($F=6.67$, $p<.01$) among patients who reported staying the same or improving in these domains. Similarly, significant differences were observed in Satisfaction with Social Roles and Activities ($F=5.86$, $p<.01$) between patients who reported worsening at six months and those who reported staying the same or improving in this domain.

Emotional Well-Being: Of the five planned comparisons [Depression, Anxiety, Emotional and Behavioral Dyscontrol, Stigma, Positive Affect and Well-being] four were statistically significant, all in the predicted direction. Specifically, statistically significant differences were observed between patients who reported worse Anxiety at six months with those who reported the same levels in this domain ($F=3.42$; $p<.05$). Similarly, significant differences were observed in Depression ($F=13.53$, $p<.01$), Stigma ($F=6.88$, $p<.01$) and Positive Affect and Well-being ($F=6.35$, $p<.01$) between patients who reported worsening at six months and those who reported staying the same or improving in these domains.

Cognitive Well-Being: Of the two planned comparisons [Applied Cognition – General Concerns, Applied Cognition – Executive Function] neither short form exhibited statistically significant changes or trends toward significance over time.

Symptomatic Well-Being: Of the five planned comparisons [Fatigue, Sleep Disturbance, Emotional and Behavioral Dyscontrol, Depression, Anxiety] one was statistically significant in the predicted direction. Specifically, differences were observed in Sleep Disturbance at six months between patients who reported worsening, staying the same and improving in this domain ($F=3.49$; $p<.05$).

Overall Quality of Life: Of the thirteen planned comparisons [all Neuro-QOL short forms] one exhibited a trend toward significance, and four were statistically significant, all in the predicted direction. Specifically, a trend toward statistical significance was observed between patients who reported worse Sleep Disturbance at six months with

those who reported staying the same or improving in these domains (F=5.45, p<.01). In addition, statistically significant differences were observed between patients who reported worse Depression (F=8.28, p<.01), Stigma (F=4.44, p<.01), Positive Affect and Well-being (F=2.98, p=.06) and Lower Extremity Function – Mobility (F=4.02, p=.02) at six months with those who reported staying the same or improving in these domains.

4.8.1.6 Conclusions. The validity of the Neuro-QOL measures for adults with stroke is supported with good internal consistency, test-retest reliability and significant correlations with many external validity measures. All Neuro-QOL short forms except Applied Cognition (Executive Function and General Concerns) were responsive to self-reported change in conceptually-related aspects of well-being.

4.8.2 RESULTS - ALS SAMPLE

4.8.2.1 Sample characteristics. Participants (N=80) were primarily male (65%), white (94%), and non-Hispanic (98%) with average age=59 years (SD=12.3). Seventy-six percent were married, 46% had a college or advanced degree. Thirty-one percent were retired, 33% on disability, 17% were employed full- and 6% were employed part time. Average time since diagnosis was 2.0 years (SD=3.6). The mean ALSFRS-R score = 32.0 (SD=8.6) with range = 8-48.

Mean T-Scores and standard deviations on the short forms are shown in Table 6. ALS patients reported significantly worse physical and social function compared to a general population reference group but similar cognitive function and more positive affect. When compared to a clinical neurological reference group, they showed greater stigma, less sleep disturbance, fatigue, depression, and emotional and behavioral dyscontrol and similar anxiety.

***4.8.2.2 Reliability*.** Internal consistency and 1 week test-retest reliability of the short forms is shown in Table 33. Cronbach's alphas range from .80 to .96 and ICCs from .49 to .93.

Table 33. Descriptive and reliability statistics for Neuro-QOL short form T-scores

Neuro-QOL Short Form	N_{items}	$N_{subjects}$	M_{GPT}	M_{CT}	*SD*	α	T-R ICCs**
Positive Affect & Well Being*	9	76	53.9		7.7	.94	.59
Applied Cognition – General Concerns*	8	77	51.8		7.1	.86	.72
Applied Cognition – Executive Function*	8	77	51.7		7.7	.84	.64
Lower Extremity Function (Mobility)*	8	57	37.6		9.9	.94	.93
Upper Extremity Function (Fine Motor, ADL)*	8	77	30.8		11.6	.96	.87
Ability to Participate in Social Roles and Activities*	8	77	42.6		7.1	.89	.71
Satisfaction with Social Roles and Activities*	8	77	42.3		5.0	.86	.49
Depression	8	77	46.6		6.4	.93	.72
Anxiety	8	77	51.5		5.4	.88	.67
Stigma	8	77		53.3	6.5	.85	.78
Fatigue	8	77		47.3	8.2	.94	.87
Sleep Disturbance	8	77		46.7	7.9	.80	.75
Emotional and Behavioral Dyscontrol	8	75		45.8	8.1	.90	.75

*For these banks, a high score indicates better function; for all other banks a high score indicates worse function
**Time 1 (baseline) vs. Time 2 (7 days)
M_{GPT} – Mean General Population T-Score; M_{CT}- Mean Clinical T-Score

***4.8.2.3 Validity*.** Table 34 shows Spearman rho correlations between Neuro-QOL short form T-scores and ALS specific measures. Table 35 presents Spearman rho correlations between Neuro-QOL short form T-Scores and cross-disease measures.

Table 34. Correlations for Neuro-QOL short form T-scores with ALS-specific measures

	ALSAQ							ALSFRS-R			
	Symbol Digit Modalities	ADL	Communication	Emotional functioning	Eating & drinking	Physical Mobility	Total	Bulbar	Fine Motor	Gross Motor	Respiratory
Depression	-0.01	0.03	0.04	.76***	0.04	0.23	0.21	0.09	0.13	0.18	0.15
Anxiety	0.08	0.14	-0.04	.53***	0.04	0.24	0.09	0.04	-0.02	0.02	0.21
Stigma	0.03	0.2	.42***	.51***	.37**	0.11	-0.17	-0.34	-0.2	0	0.06
Positive Affect & Well-being	0.11	0	0.04	-.66***	0.05	-0.18	-0.21	-0.11	-0.22	-0.12	0.04
Applied Cognition-General Concerns	.51***	-0.1	-0.2	-.36**	-0.24	0.01	-0.02	0.1	-0.06	-0.14	-0.03
Applied Cognition – Executive Functioning	.51***	-0.17	-0.18	-0.17	-0.28	0.05	0.08	0.17	0.1	-0.09	0.05
Lower Extremity Function - Mobility	0.05	-.67***	-0.05	-0.34	0	-.65***	0.33	-0.04	0.34	.66***	0.07
Upper Extremity Function - Fine motor, ADL	0.15	-.88***	-0.21	-0.14	-0.25	-.43***	.66***	0.24	.79***	.54***	0.13
Ability to participate in social roles & activities	0.1	-.55***	-0.19	-.44***	-0.09	-.41***	.30*	0.07	0.28	.31*	0.13
Satisfaction with social roles & activities	0.16	-.43***	-0.18	-.50***	-0.07	-.52***	0.24	0.07	0.21	.30*	0.13
Fatigue	0	0.06	0.13	.49***	0.16	0.06	0.1	-0.03	0.11	0.15	0.01
Sleep Disturbance	-0.24	0.12	0.14	.35*	0.24	0	0.03	-0.11	0.04	0.21	0.04
Emotional & Behavioral Dyscontrol	0.01	0.23	-0.06	.34*	-0.11	.37**	-0.03	0.03	-0.12	0.1	0.13

*p < .05; **p < .01; ***p <.001

Table 35. Correlations for Neuro-QOL short form T-scores with cross-disease measures

Neuro-QOL Short Form	Barthel Index	Lawton IADL Scale	KPSS	EQ-5D Index Score	PROMIS Mental Health T-Score	PROMIS Physical Function T-Score	Global HRQL (0-4)	Pain Scale (0-10)
Depression	.08	-.06	.004	-.18	-.67***	-.31**	-.53***	.27*
Anxiety	-.07	-.14	-.15	-.29	-.49***	-.35**	-.33**	.29*
Stigma	-.15	-.22	-.08	-.28	-.39***	-.25*	-.08	.16
Positive Affect & Well Being	-.14	.07	-.05	.12	.68***	.32**	.55***	-.22
Applied Cognition – General Concerns	.03	-.13	.09	.17	.29	.11	.13	-.38***
Applied Cognition – Executive Function	.07	.08	.17	.17	.21	.07	.07	-.15
Lower Extremity (Mobility)	.64***	.54***	.55***	.59***	.27	.66***	.16	.10
Upper Extremity (Fine Motor, ADL)	.76***	.58***	.7***	.69***	.14	.37	.02	.03
Ability to Participate in Social Roles and Activities	.38***	.42***	.47***	.51***	.48***	.63***	.47***	-.15
Satisfaction with Social Roles and Activities	.40***	.41***	.41***	.48***	.47***	.63***	.36**	-.23*
Fatigue	.14	-.04	-.05	-.02	-.46***	-.32**	-.34**	.20
Sleep Disturbance	.04	.05	-.1	-.12	-.4***	-.22	-.26*	.44***
Emotional and Behavioral Dyscontrol	-.12	-.13	-.16	-.28	-.37**	-.24*	-.23*	.26*

*p < .05; **p < .01; ***p < .001

4.8.2.4 Known Groups Validity. In the baseline assessment, the extent to which ALS patients agreed with the statement "I am content with my quality of life right now" was significantly associated with the following Neuro-QOL short forms: Depression, Anxiety, Positive psychological functioning, Social role - participation, Social role - satisfaction, and Fatigue. The corresponding effect sizes ranged from .22 to 2.86.

***4.8.2.5 Responsiveness*.** Of the 32 planned comparisons, 4 were statistically significant and 1 exhibited a trend toward significance, all in the predicted direction.

Physical Well-being: Of the four planned comparisons, one was significant. Specifically, patients who reported a worsening of their physical well-being showed significantly worse Upper Extremity Function scores than those who reported no change ($t=2.17$; $p<.05$).

Cognitive Well-being: Of the two planned comparisons, one was significant. Patients with worsening cognitive well-being reported significantly worsening executive function compared to those who did not have a change in cognitive well-being ($t=3.22$; $p<.01$).

Emotional Well-being: Of the five planned comparisons, one was significant. Patients who reported decreasing emotional well-being showed increased scores on the Depression Short Form (F=3.30; p<.05).

Social/Family Well-being: Of the three planned comparisons, none were significant.

Symptomatic Well-being: Of the five planned comparisons, none were significant.

Overall Quality of Life: Of the thirteen planned comparisons, one was significant and one approached significance. Specifically, patients who reported a decrease in overall quality of life also showed significant worsening of upper extremity function (t=3.17; p<.05) and a trend toward increasing fatigue (t=-1.68; p<.10

4.8.2.6 Conclusions. The study sample represented a wide range of functioning, similar to an ALS clinic population. Internal consistency was high for 11, and adequate for 2, of the 13 Neuro-QOL scales. The Intraclass Correlation Coefficients (ICC) ranged from .49 (satisfaction with social roles) to .94 (mobility), suggesting that further evaluation of test-retest reliability is warranted in some cases. Convergent and concurrent validity are high, with correlations of the expected strength and in the expected direction. Several Neuro-QOL short forms (Upper Extremity Function, Applied Cognitive –Executive Function, and Depression) demonstrated responsiveness to self-reported change. The remaining short forms did not.

4.8.3 RESULTS – MS SAMPLE

***4.8.3.1 Sample Characteristics*.** Participants (N=161) were primarily female (86%), white (88%), and non-Hispanic (93%) with average age=49.8 years (SD=10.5). 58.4% were married, 90% had some college or a college degree. Thirty-seven percent were on disability and 34% were employed full time. MSFC scores ranged from -2.90 to 1.7, with mean=0.0 (SD=.69). Mean MS Performance Scale score = 16.04 (SD=9.18; range = 0-35).

Mean T-Scores and standard deviations on the short forms are shown in Table 9. MS patients reported worse physical, social and cognitive function compared to a general population reference group but greater positive affect. When compared to a clinical neurological reference group, they showed less depression and better emotional and behavioral control but similar levels of stigma, sleep disturbance, fatigue and anxiety.

***4.8.3.2 Reliability*.** Internal consistency and 1 week test-retest reliability of the short forms is shown in Table 36. Cronbach's alphas range from .81 to .95 and ICCs from .67 to .89.

Table 36. Descriptive and reliability statistics for Neuro-QOL short form T-scores

Neuro-QOL Short Form	N_{items}	$N_{persons}$	M_{GPT}	M_{CT}	*SD*	α	T-R ICCs**
Positive Affect & Well Being*	9	161	53.61		7.72	.95	.76
Applied Cognition – General Concerns*	8	161	42.56		8.70	.95	.83
Applied Cognition – Executive Function*	8	161	46.02		9.37	.90	.86
Lower Extremity (Mobility)*	8	149	43.55		9.44	.93	.89
Upper Extremity (Fine Motor, ADL)*	8	161	44.03		9.21	.86	.81
Ability to Participate in Social Roles and Activities*	8	161	46.02		7.43	.95	.73
Satisfaction with Social Roles and Activities*	8	161	44.97		6.07	.89	.76
Depression	8	161	46.69		6.93	.93	.68
Anxiety	8	161	51.32		6.88	.93	.67
Stigma	8	161		49.35	7.23	.86	.69
Fatigue	8	161		48.81	8.52	.95	.80
Sleep Disturbance	8	161		48.50	8.60	.81	.77
Emotional and Behavioral Dyscontrol	8	161		46.78	8.63	.91	.74

*For these banks, a high score indicates better function; for all other banks a high score indicates worse function
**Time 1 (baseline) vs. Time 2 (7 days)
M_{GPT} – Mean General Population T-Score; M_{CT}- Mean Clinical T-Score

4.8.3.3 Validity: Table 37 shows Spearman rho correlations between Neuro-QOL short form T-scores and MS specific measures. Table 38 presents Spearman rho correlations between Neuro-QOL short form T-Scores and cross-disease measures.

Table 37. Correlations for Neuro-QOL short form T-scores with MS-specific measures

Neuro-QOL Short Form	FAMS	FAMS Mobility	FAMS Symptoms	FAMS Emotional Well-Being	FAMS General Contentment	FAMS Thinking and Fatigue	FAMS Family/Social Well-Being	FAMS Additional Concerns	MS Functional Composite	The MS Performance Scales
Depression	-.71***	-.41***	-.48***	-.76***	-.72***	-.57***	-.58***	-.63***	-0.15	.48***
Anxiety	-.60***	-.28***	-.43***	-.62***	-.57***	-.60***	-.49***	-.58***	-0.09	.32***
Stigma	-.77***	-.71***	-.44***	-.70***	-.66***	-.54***	-.60***	-.60***	-.37***	.66***
Positive Affect & Well Being	.77***	.50***	.45***	.78***	.86***	.58***	.60***	.67***	.16*	-.50***
Applied Cognition – General Concerns	.63***	.35***	.48***	.38***	.46***	.77***	.52***	.54***	.21**	-.57***
Applied Cognition – Executive Function	.61***	.38***	.44***	.42***	.46***	.69***	.48***	.49***	.32***	-.58***
Lower Extremity Function - Mobility	.59***	.86***	.46***	.44***	.41***	.35***	.23***	.46***	.55***	-.75***
Upper Extremity Function -Fine Motor, ADL	.58***	.66***	.42***	.45***	.44***	.45***	.30***	.46***	.59***	-.73***
Ability to Participate in Social Roles and Activities	.81***	.71***	.57***	.67***	.73***	.66***	.54***	.65***	.24**	-.68***
Satisfaction with Social Roles and Activities	.83***	.72***	.55***	.72***	.72***	.66***	.58***	.63***	.32***	-.71***
Fatigue	-.81***	-.52***	-.67***	-.63***	-.67***	-.84***	-.58***	-.64***	-.17*	.63***
Sleep Disturbance	-.67***	-.32***	-.56***	-.60***	-.62***	-.69***	-.53***	-.62***	-0.03	.44***
Emotional and Behavioral Dyscontrol	-.60***	-.32***	-.45***	-.51***	-.47***	-.65***	-.52***	-.61***	-.21**	.45***

*p < .05; **p < .01; ***p < .001

Table 38. Correlations for Neuro-QOL short form T-scores with cross-disease measures

Neuro-QOL Short Form	Barthel Index	Karnofsky Performance Scale	Lawton IADL Scale	Symbol Digit Modalities # Correct	Symbol Search Raw Score	Digit Symbol Coding # Correct	PROMIS Physical Function T- Score	PROMIS Mental Health T-Score	Pain Scale (0-10)	EQ-5D Index Score	Global HRQL (0-4)
Depression	-.23**	-.28***	-.27***	-0.05	-0.1	-.20*	-.54***	-.75***	.42***	-.46***	-.66***
Anxiety	-0.07	-0.15	-.20*	-0.05	-0.04	-0.1	-.46***	-.69***	.35***	-.40***	-.52***
Stigma	-.45***	-.59***	-.43***	-.17*	-.22**	-.29***	-.63***	-.60***	.42***	-.56***	-.54***
Positive Affect & Well Being	.22**	.28***	.27***	0.01	0.05	0.12	.61***	.81***	-.40***	.48***	.81***
Applied Cognition – General Concerns	.19*	.23**	.29***	.23**	0.14	.24**	.48***	.58***	-.38***	.49***	.42***
Applied Cognition – Executive Function	.19*	.26***	.30***	.34***	.22**	.32***	.50***	.56***	-.34***	.44***	.44***
Lower Extremity (Mobility)	.68***	.80***	.42***	.25**	.38***	.50***	.65***	.31***	-.49***	.65***	.35***
Upper Extremity (Fine Motor, ADL)	.59***	.62***	.51***	.33***	.40***	.53***	.65***	.42***	-.43***	.60***	.36***
Ability to Participate in Social Roles and Activities	.41***	.45***	.39***	0.09	0.14	.24**	.77***	.69***	-.49***	.59***	.71***
Satisfaction with Social Roles and Activities	.47***	.51***	.41***	0.13	.17*	.28***	.73***	.68***	-.50***	.62***	.68***
Fatigue	-.23**	-.28***	-.30***	-0.05	-0.05	-0.12	-.72***	-.69***	.46***	-.52***	-.62***
Sleep Disturbance	-0.14	-.19*	-.16*	-0.01	-0.04	-0.08	-.59***	-.69***	.44***	-.44***	-.57***
Emotional and Behavioral Dyscontrol	-.16*	-.27***	-.27***	-0.11	-0.06	-0.11	-.47***	-.62***	.35***	-.41***	-.44***

*p = .05; **p = .01; ***p = .001

4.8.3.4 Known Groups Validity. Patients grouped according to MSFC quartile scored significantly differently on all Neuro-QOL SFs, except Anxiety, Depression, and Emotional & Behavioral Dyscontrol, with effect sizes ranging from .47 to 2.15.

***4.8.3.5 Responsiveness*.** Of the 32 planned comparisons, 18 were statistically significant and 3 exhibited a trend toward significance, in the predicted direction.

Physical Well-being: Of the four planned comparisons, one was significant and one exhibited a trend toward significance, both in the predicted direction. Specifically, patients who reported a worsening of their physical well-being showed worsening of scores on Physical Function – Lower Extremity (extended assessment; $F=4.36$; $p<.05$) and a trend toward worse fatigue ($F=2.36$; $p<.10$).

Cognitive Well-being: Of the two planned comparisons, both were significant and in the predicted direction. Patients who reported worsening cognitive well-being showed worsening of their cognitive function, both in terms of general concerns ($F=7.09$; $p<.01$) and executive function ($F=4.69$; $p<.01$).

Emotional Well-being: Of the five planned comparisons, four were significant and one showed a trend toward significance in the predicted direction. Patients who reported worsening emotional well-being also reported increased depression ($F=14.82$; $p<.0001$), anxiety ($F=7.28$; $p<.01$) and emotional and behavioral dyscontrol ($F=3.19$; $p<.05$) and decreased positive affect and well-being. Patients who reported increased emotional well-being showed a trend toward scoring lower on the Stigma Short Form ($F=2.61$; $p<.10$).

Social/Family Well-being: Of the three planned comparisons, one was significant. Specifically, patients who reported improved social/family well-being at 6 months also reported decreasing stigma ($F=3.21$, $p<.05$).

Symptomatic Well-being: Of the five planned comparisons, three were significant. Patients who reported worsened symptomatic well-being showed worsening on the Depression Short Form ($F=5.02$; $p<.01$). Patients who reported improved symptomatic well-being showed decreased fatigue ($F=6.45$; $p<.01$) and improved emotional and behavioral control ($F=3.14$; $p<.05$).

Overall Quality of Life: Of the thirteen planned comparisons, seven were significant and one showed a trend toward significance. Patients who reported decreased overall quality of life also showed worsening depression ($F=8.99$; $p<.001$), anxiety ($F=5.57$; $p<.05$), ability to participate in social roles and activities ($F=3.91$; $p<.05$) and a trend toward decreased upper extremity function ($F=2.51$; $p<.10$).

***4.8.3.6 Conclusions*.** The study sample was generally representative of MS clinic populations. The 13 Neuro-QOL scales demonstrated high internal consistency. The Intraclass Correlation Coefficients (ICC) were acceptable, ranging from .67 (anxiety) to .89 (lower extremity).Convergent validity with generic and legacy measures was good; correlations were of the expected strength and direction and short forms discriminated between patients grouped according to disease severity.There is some initial evidence for Neuro-QOL short form responsiveness to self-reported change in MS patients, particularly for the short forms assessing emotional and cognitive well-being, where 4 of 5 and 2 of 2 planned comparisons were significant.

4.8.4 RESULTS – PD SAMPLE

4.8.4.1 Sample Characteristics. Participants were primarily male (62%), white (95%), and non-Hispanic (97%) with average age=65. Seventy-four percent were married, 55% had a college or advanced degree. Fifty-eight percent were retired and 20% were employed either full or part time. Most (76%) were in mild stages of the disease: Hoehn and Yahr 1 (N=19; 16%), 2 (N=72; 60%), 3 (N=23; 19%), 4 (N=6; 5%). Average time since PD diagnosis was 7.1 years. 80% were taking L-Dopa either alone or in combination with other anti-PD medications and 9% reported undergoing prior PD surgery. A majority of patients (55%) were primarily affected on their right side; most experienced no (43%) or little (33%) activity limitation due to motor fluctuations.

Mean T-Scores and standard deviations on the Neuro-QOL short forms are shown in Table 12. PD patients reported worse cognitive, physical and social function compared to a general population reference group but more positive affect and well-being. When compared to a clinical neurological population, they showed less sleep disturbance, fatigue and depression and a greater sense of emotional and behavioral control.

***4.8.4.2 Reliability*.** Internal consistency and 1 week test-retest reliability of the short forms is shown in Table 39. Cronbach's alphas range from .82 to .94 and ICCs from .80 to .89.

Table 39. Descriptive and reliability statistics for Neuro-QOL short form T-scores

Neuro-QOL Short Form	N_{items}	$N_{persons}$	M_{GPT}	M_{CT}	*SD*	α	T-R ICCs**
Positive Affect & Well Being*	9	120	54.40		7.53	.94	.86
Applied Cognition – General Concerns*	8	120	44.35		7.62	.90	.84
Applied Cognition – Executive Function*	8	120	46.25		8.38	.90	.87
Lower Extremity Function (Mobility)*	8	118	45.80		7.54	.84	.88
Upper Extremity Function (Fine Motor, ADL)*	8	120	42.28		8.34	.82	.84
Ability to Participate in Social Roles and Activities*	8	120	47.85		6.83	.94	.83
Satisfaction with Social Roles and Activities*	8	119	46.21		5.70	.89	.80
Depression	8	119	45.85		6.86	.93	.81
Anxiety	8	120	50.82		6.80	.91	.87
Stigma	8	120		48.39	6.62	.85	.87
Fatigue	8	119		46.04	7.75	.93	.88
Sleep Disturbance	8	120		47.70	7.98	.81	.89
Emotional and Behavioral Dyscontrol	8	120		43.49	8.36	.91	.84

*For these banks, a high score indicates better function; for all other banks a high score indicates worse function
**Time 1 (baseline) vs. Time 2 (7 days); M_{GPT} – Mean General Population T-Score; M_{CT}- Mean Clinical T-Score

***4.8.4.3 Validity*.** Spearman rho correlations between the Neuro-QOL short forms and the PD-specific measures are shown in Table 40 and between the Neuro-QOL short forms and the cross-disease instruments in Table 41.

Table 40. Correlations for Neuro-QOL short form T-scores with PD-specific measures

	PDQ-39								UPDRS****				Mo CA Total	PHQ-9 Total
Neuro-QOL Short Form	Mobility	ADL	EWB	Stigma	Social support	CI	Comm	BD	Total	Part 1	Part 2	Part 3		
Positive Affect & Well Being	-.48***	-.36***	-.56***	-0.17	-.45***	-.41***	-.44***	-0.18	-.29***	-.30***	-.27**	-0.07	0.17	-.50***
Applied Cognition – General Concerns	-.34***	-.35***	-.23*	-0.17	-.42***	-.49***	-.42***	-.25**	-0.18	-.29***	-.23**	-.24**	.20*	-.32***
Applied Cognition – Executive Function	-.44***	-.37***	-.34***	-0.07	-.35***	-.51***	-.42***	-.23*	-.31***	-.26**	-.32***	-0.14	.37***	-.24**
Lower Extremity Function (Mobility)	-.72***	-.61***	-.36***	-.23*	-.32***	-.38***	-.41***	-.38***	-.58***	-.22*	-.59***	-0.14	0.04	-.33***
Upper Extremity Function (Fine Motor, ADL)	-.46***	-.76***	-.37***	-.35***	-.40***	-.42***	-.41***	-.24**	-.34***	-0.14	-.44***	-0.11	0.09	-.27**
Ability to Participate in Social Roles and Activities	-.69***	-.46***	-.43***	-.24**	-.44***	-.43***	-.55***	-.36***	-.37***	-.37***	-.41***	-0.13	.21*	-.50***
Satisfaction with Social Roles and Activities	-.62***	-.48***	-.51***	-.29***	-.52***	-.38***	-.50***	-.31***	-.39***	-.30***	-.46***	-.23*	.25**	-.55***
Depression	.38***	.36***	.68***	.19*	.36***	.33***	.35***	0.18	.21*	.32***	.21*	0.02	-0.13	.47***
Anxiety	.39***	.40***	.70***	.38***	.28**	.41***	.30***	.24**	.22*	.35***	.20*	0.03	-0.06	.42***
Stigma	.49***	.46***	.51***	.52***	.44***	.34***	.45***	.40***	.19*	0.18	.28**	0.18	-.20*	.46***
Fatigue	.67***	.47***	.56***	.36***	.39***	.53***	.54***	.54***	.35***	.28**	.39***	.20*	-0.17	.63***
Sleep Disturbance	.47***	.47***	.47***	.39***	.35***	.54***	.46***	.46***	.24**	.31***	.32***	.21*	-0.14	.54***
Emotional & Behavioral Dyscontrol	.35***	.45***	.49***	.27**	.46***	.40***	.33***	.20*	0.12	.22*	.18*	0.05	-0.17	.33***

*p = .05; **p = .01; ***p = .001

**** Non-standard scoring was used for UPDRS Part 3

EWB=Emotional Well-being; CI=Cognitive Impairment; Comm=Communication

Table 41. Correlations for Neuro-QOL short form T-scores with cross-disease measures

Neuro-QOL Short Form	Barthel Index	Lawton IADL Scale	Oral Symbol Digit Modalities # Correct	Symbol Search Raw Score	Digit Symbol Coding # Correct	PROMIS Global Physical	PROMIS Global Mental	EQ-5D Index Score	Global HRQL (0-4)
Positive Affect & Well Being	.24**	.17	.16	.20*	.13	.45***	.74***	.41***	.64***
Applied Cognition – General Concerns	.25**	.05	.24**	.15	.11	.30***	.41***	.18	.27**
Applied Cognition – Executive Function	.35***	.28**	.41***	.32***	.34***	.39***	.39***	.21*	.29***
Lower Extremity (Mobility)	.51***	.07	.10	.02	.05	.55***	.35***	.57***	.23*
Upper Extremity (Fine Motor, ADL)	.46***	.27**	.11	.03	.02	.39***	.37***	.41***	.29***
Ability to Participate in Social Roles and Activities	.26**	.11	.20*	.23*	.16	.55***	.64***	.44***	.52***
Satisfaction with Social Roles and Activities	.31***	.18	.15	.19	.17	.46***	.64***	.45***	.53***
Depression	-.30***	-.12	-.16	-.09	.001	-.36***	-.65***	-.41***	-.54***
Anxiety	-.37***	-.12	-.12	-.06	-.01	-.45***	-.61***	-.42***	-.45***
Stigma	-.33***	-.14	-.02	-.03	-.51***	-.42***	-.51***	-.38***	-.43***
Fatigue	-.35***	.02	-.06	-.08	-.005	-.62***	-.53***	-.44***	-.39***
Sleep Disturbance	-.26**	-.07	-.06	-.01	.01	-.48***	-.44***	-.32***	-.28**
Emotional and Behavioral Dyscontrol	-.28**	-.12	-.11	-.004	.10	-.35***	-.38***	-.30***	-.27**

*p ≤ .05; **p ≤ .01; ***p ≤ .001

***4.8.4.4 Known Groups Validity*.** Patients in H & Y Stage 1 or 2 scored significantly differently on all Neuro-QOL SFs, except Applied Cognition-General Concerns and Emotional & Behavioral Dyscontrol, than did patients in Stages 3 or 4, with effect sizes ranging from .5 to 1.11.

***4.8.4.5 Responsiveness*.** Of the 32 planned comparisons, 7 were statistically significant and 1 exhibited a trend toward significance, in the predicted direction.

Physical Well-being: Of the four planned comparisons, two were significant in the predicted direction. Specifically, patients who reported a worsening of their physical well-being showed worsening of scores on Fatigue (F=8.13; p<.01) Lower Extremity Function (extended assessment; F=4.69; p< .05).

Cognitive Well-being: Of the two planned comparisons, none were significant.

Emotional Well-being: Of the five planned comparisons, one showed a trend toward significance. Patients who reported changes in emotional well-being also exhibited a trend toward having changes in positive affect and well-being.

Social/Family Well-being: Of the three planned comparisons, none were significant.

Symptomatic Well-being: Of the five planned comparisons, one was significant. Specifically, patients who reported worsening symptomatic well-being also demonstrated worsening scores on Fatigue (extended assessment; $F=3.32$; $p<.05$).

Overall Quality of Life: Of the thirteen planned comparisons, four were significant. Patients who reported a worsening of overall quality of life showed decreasing positive affect and well-being ($F=6.73$; $p<.01$), ability to participate in social activities ($F=4.04$; $p<.05$), and upper extremity function ($F=5.33$; $p<.01$) and increasing fatigue (extended assessment, $F=3.63$; $p<.05$).

4.8.4.6 Conclusions. The Neuro-QOL measures demonstrated high internal consistency and test-retest reliability. Convergent validity was supported by correlations with generic and PD-specific measures in the expected directions. Correlations were generally modest in strength, warranting additional validation in PD samples. Neuro-QOL measures showed good discrimination between patients at different levels of disease severity. There was only limited evidence for responsiveness to self-reported changes in different domains of well-being.

4.8.5 RESULTS - ADULT EPILEPSY

4.8.5.1 Sample Characteristics. Participants were primarily male (51%), white (85%), and non-Hispanic (75%) with average age=47.3 (Range = 18-93). Forty-seven percent were married, 67% had some college or beyond. Fourteen percent were retired, 22% on disability and 37% were employed either full or part time. Average time since epilepsy diagnosis was 18.5 years (SD=13.9). Generalized seizures were most frequently experienced (57%) followed by focal seizures (25%). Mean number of seizures in the past 3 months = 10.7 (SD=37.6). 95% were taking medication for their seizure disorder, with 64% of those on polytherapy. Twelve percent had undergone surgery for their epilepsy.

Mean T-Scores and standard deviations on the short forms are shown in Table 15. Epilepsy patients reported significantly worse cognitive and social function compared to a general population reference group but similar levels of physical function and greater positive affect and well-being. When compared to a clinical neurological population, they showed similar levels of stigma, greater anxiety, but less depression, sleep disturbance, fatigue, and sense of emotional and behavioral dyscontrol.

***4.8.5.2 Reliability*.** Internal consistency and 1 week test-retest reliability of the short forms is shown in Table 42. Cronbach's alphas range from .86 to .96 and ICCs from .57 to .89.

Table 42. Descriptive and reliability statistics for Neuro-QOL short form T-scores

Neuro-QOL Short Form	N_{items}	$N_{persons}$	M_{GPT}	M_{CT}	*SD*	α	T-R ICCs**
Positive Affect & Well Being*	9	118	53.8		8.2	0.95	0.81
Applied Cognition – General Concerns*	8	119	41.9		8.7	0.94	0.82
Applied Cognition – Executive Function*	8	119	43.6		10.3	0.94	0.87
Lower Extremity Function -Mobility*	8	114	50.4		9.0	0.92	0.89
Upper Extremity Function -Fine Motor, ADL*	8	119	49.0		7.7	0.88	0.87
Ability to Participate in Social Roles and Activities*	8	119	45.3		7.2	0.94	0.57
Satisfaction with Social Roles and Activities*	8	119	45.9		6.5	0.89	0.72
Depression	8	118		47.9	8.3	0.96	0.82
Anxiety	8	118		52.3	8.1	0.94	0.81
Stigma	8	119		49.7	9.1	0.91	0.83
Fatigue	8	119		45.6	9.4	0.95	0.81
Sleep Disturbance	8	119		48.2	9.8	0.86	0.77
Emotional and Behavioral Dyscontrol	8	119		46.3	10.1	0.93	0.84

* For these banks, a high score indicates better function; for all other banks a high score indicates worse function; ** Time 1 (baseline) vs. Time 2 (7 days); M_{GPT} – Mean General Population T-Score; M_{CT}- Mean Clinical T-Score

***4.8.5.3 Validity*.** Spearman correlations between Neuro-QOL short forms and epilepsy-specific and cross-disease measures are shown in Tables 43 and 44.

Table 43. Correlations for Neuro-QOL short form T-scores with epilepsy-specific measures

Neuro-QOL Short Form	QOLIE-31								Liverpool Seizure Severity Scale	Liverpool Adverse Events Profile
	Total	Cognitive	Energy/ Fatigue	Emotional Well-Being	Medication Effects	Overall Quality of Life	Social Function	Seizure Worry		
Positive Affect & Well Being	.737 **	.522 **	.543 **	.671 **	.423 **	.617 **	.643 **	.520 **	-.361 **	-.563 **
Applied Cognition – General Concerns	.677 **	.784 **	.534 **	.428 **	.428 **	.422 **	.394 **	.401 **	-0.188	-.699 **
Applied Cognition – Executive Function	.572 **	.668 **	.395 **	.415 **	.260 **	.411 **	.351 **	.247 **	0.005	-.511 **
Lower Extremity Function - Mobility	.330 **	.338 **	.280 **	0.183	.213 *	0.168	.249 **	.212 *	-0.198	-.393 **
Upper Extremity Function - Fine Motor, ADL	.334 **	.281 **	.271 **	.205 *	0.123	.210 *	.299 **	.232 *	-0.207	-.355 **
Ability to Participate in Social Roles and Activities	.646 **	.486 **	.466 **	.536 **	.419 **	.458 **	.599 **	.427 **	-.307 *	-.523 **
Satisfaction with Social Roles and Activities	.544 **	.386 **	.472 **	.464 **	.316 **	.383 **	.487 **	.409 **	-0.22	-.340 **
Depression	-.642 **	-.430 **	-.520 **	-.699 **	-.310 **	-.573 **	-.524 **	-.438 **	.386 **	.451 **
Anxiety	-.617 **	-.421 **	-.526 **	-.690 **	-.352 **	-.453 **	-.476 **	-.550 **	.442 **	.482 **
Stigma	-.582 **	-.365 **	-.419 **	-.504 **	-.373 **	-.420 **	-.574 **	-.501 **	.407 **	.484 **
Fatigue	-.584 **	-.405 **	-.665 **	-.441 **	-.381 **	-.299 **	-.500 **	-.510 **	.487 **	.610 **
Sleep Disturbance	-.528 **	-.413 **	-.460 **	-.421 **	-.367 **	-.329 **	-.428 **	-.471 **	.380 **	.634 **
Emotional and Behavioral Dyscontrol	-.579 **	-.479 **	-.453 **	-.539 **	-.342 **	-.386 **	-.483 **	-.393 **	.332 *	.553 **

*p < .05; **p < .01

Table 44. Spearman's Rho Correlations for Neuro-QOL short form T-scores with cross-disease measures

Neuro-QOL Short Form	Barthel Index	Lawton IADL Scale	Symbol Digit Modalities # Correct	Symbol Search Raw Score	Digit Symbol Coding # Correct	PROMIS Global Physical	PROMIS Global Mental	Pain Scale 0-10	EQ-5D Index Score	Global HRQL
Positive Affect & Well Being	.185 *	.216 *	-0.088	-0.03	0.005	.480 **	.732 **	-.395 **	.486 **	.597 **
Applied Cognition – General Concerns	.264 **	.231 *	-0.092	-0.077	0.046	.523 **	.542 **	-.325 **	.425 **	.278 **
Applied Cognition – Executive Function	.308 **	.361 **	0.111	0.085	.238 *	.444 **	.453 **	-.286 **	.426 **	.201 *
Lower Extremity Function (Mobility)	.527 **	.382 **	0.15	0.126	0.169	.450 **	.283 **	-.330 **	.490 **	.215 *
Upper Extremity Function (Fine Motor, ADL)	.597 **	.442 **	0.157	0.094	.318 **	.494 **	.278 **	-.387 **	.515 **	0.172
Ability to Participate in Social Roles and Activities	.357 **	.323 **	0.03	-0.001	0.107	.493 **	.617 **	-.359 **	.495 **	.462 **
Satisfaction with Social Roles and Activities	.270 **	0.149	0.02	0.049	0.116	.457 **	.530 **	-.313 **	.427 **	.568 **
Depression	-0.02	-0.111	0.088	-0.041	-0.062	-.417 **	-.722 **	.290 **	-.407 **	-.641 **
Anxiety	-0.055	-0.075	0.063	-0.057	-0.086	-.348 **	-.561 **	.245 **	-.335 **	-.503 **
Stigma	-0.136	-.188 *	0.119	0.013	-0.059	-.371 **	-.527 **	.192 *	-.343 **	-.349 **
Fatigue	-0.16	-0.141	0.087	-0.004	-0.075	-.525 **	-.455 **	.261 **	-.357 **	-.283 **
Sleep Disturbance	-0.12	-0.105	0.128	0.113	0.082	-.423 **	-.429 **	0.172	-.337 **	-.247 **
Emotional and Behavioral Dyscontrol	-0.175	-0.155	0.169	0.082	-0.01	-.298 **	-.498 **	0.093	-.301 **	-.393 **

* =p< .05; ** = p< 0.01

***4.8.5.4 Known Groups Validity*.** Statistically significant known group differences were observed between Leeds Seizure Severity Scale quartile groups and the following Neuro-QOL short forms: Anxiety (F=5.15, p<.01), Depression (F=5.71, p<.01), Emotional and Behavioral Dyscontrol (F=4.32, p<.01), Fatigue (F=9.08, p<.01), Positive Affect and Well-being (F=6.3, p<.01), Sleep Disturbance (F=3.36, p<.01), Stigma (F=4.65, p<.01) and Upper Extremity - Fine Motor, ADL (F=4.07, p<.01).

***4.8.5.5 Responsiveness*:** Of the 32 planned comparisons, nine were statistically significant and five exhibited a trend toward significance, in the predicted direction.

Physical Well-Being: Of the four planned comparisons [Lower Extremity Function-Mobility, Upper Extremity Function - Fine Motor, ADL, Fatigue, and Sleep Disturbance] two were statistically significant and one exhibited a trend toward significance, all in the predicted direction. Specifically, a trend toward significance was observed between patients who reported worse Physical Function – Lower Extremity at six months with those who reported better functioning (F=2.74; p=.069). Statistically significant differences were observed between patients who reported worsening at six months with those who reported staying the same or improving in both Fatigue (F=4.94; p<.01) and Sleep Disturbance (F=3.21, p<.05).

Social/Family Well-Being. Of the three planned comparisons [Ability to Participate in Social Roles and Activities, Satisfaction with Social Roles and Activities, Stigma] one exhibited a trend toward significance, in the predicted direction. Specifically, a trend toward significance was observed between patients who reported

worse Ability to Participate in Social Roles and Activities at six months with those who reported improvements in this domain (F=2.64; p=.076).

Emotional Well-Being. Of the five planned comparisons [Depression, Anxiety, Emotional and Behavioral Dyscontrol, Stigma, Positive Affect and Well-being] three were statistically significant and one exhibited a trend toward significance, all in the predicted direction. Specifically, a trend toward significance was observed between patients who reported worse Anxiety at six months with those who reported improvements in this domain (F=2.62; p=.077). Statistically significant differences were observed between patients who reported worse Depression at six months with those who reported improvements in this domain (F=4.94; p<.01); between patients who reported the same level of Emotional and Behavioral Dyscontrol with those who reported improvements in this domain (F=3.19, p<.05); and between patients who reported improved Positive Affect and Well-being with those who reported staying the same or worsening in this domain (F= 7.40, p<.01).

Cognitive Well-Being. Of the two planned comparisons [Applied Cognition – General Concerns, Applied Cognition – Executive Function] neither short form exhibited statistically significant changes or trends toward significance over time.

Symptomatic Well-Being. Of the five planned comparisons [Fatigue, Sleep Disturbance, Emotional and Behavioral Dyscontrol, Depression, Anxiety] one was statistically significant in the predicted direction. Specifically, differences were observed between patients who reported worse Depression at six months with those who reported staying the same or improving in this domain (F=3.94; p<.05).

Overall Quality of Life. Of the thirteen planned comparisons [all Neuro-QOL short forms] two were statistically significant and three exhibited a trend toward significance, all in the predicted direction. Specifically, a trend toward significance was observed between patients who reported staying the same and those who reported improving in their scores of Emotional and Behavioral Dyscontrol (F=3.07, p=.051), Anxiety (F=2.97, p=.056), Fatigue (F=2.92, p=.058), and Ability to Participate in Social Roles and Activities (F=2.86, p=.061). Statistically significant differences were observed between patients who reported worse Depression over time with those who reported staying the same or improving in this domain (F=3.71; p<.05). Significant differences were also observed between patients who reported improvements in Positive Affect and Well-being at six months compared to those who reported staying the same or worsening in this domain (F=6.39, p<.01).

4.8.5.6 **Conclusions.** The 13 Neuro-QOL scales demonstrated high internal consistency, ranging from .86 (Sleep disturbance) to .96 (Depression). The Intraclass Correlation Coefficients (ICC) were generally acceptable, ranging from .57 (Ability to Participate in Social Roles and Activities) to .89 (Lower Extremity Function – Mobility). Convergent and discriminant validity were good, with correlations of the expected strength and in the expected direction. Neuro-QOL measures discriminated between patients at different levels of disease severity. There is initial evidence of responsiveness. Self-reported changes in physical, emotional and symptomatic well-being and overall quality of life were reflected in significant changes in conceptually-related Neuro-QOL short forms.

4.8.6 RESULTS - PEDIATRIC EPILEPSY

4.8.6.1 Sample Characteristics. Participants (N=61) were primarily male (62.3%), white (75.9%), and non-Hispanic (75.4%) with average age=13.4 (SD=2.6; range = 10 to 18). At baseline, 17.8% reported having seizures daily, 13.3% weekly, 35.6% monthly and 33.3% yearly, and all patients were taking anti-epilepsy drugs at the time of testing.

Mean T-Scores and standard deviations on the short forms are shown in Table 18. Pediatric epilepsy patients reported better function/less symptoms on all domains compared to the reference group.

4.8.6.2 Reliability. Internal consistency and 1 week test-retest reliability of the short forms is shown in Table 45. Cronbach's alphas range from .76 to .87 and ICCs from .44 to .94.

Table 45. Descriptive and reliability statistics for Neuro-QOL short form T-scores

Neuro-QOL Short Form	N_{items}	$N_{persons}$	M_{GPT}	M_{CT}	*SD*	α	T-R** ICCs
Social Relations – Interactions with Peers*	8	59	52.70		9.77	.86	.58
Applied Cognition – General Concerns*	8	61		52.29	7.20	.86	.69
Depression	8	59	45.16		7.13	.85	.69
Anxiety	8	58	49.02		7.58	.76	.67
Stigma	8	61		45.39	5.73	.79	.44
Fatigue	8	61		48.42	7.75	.80	.52
Pain	10	59		46.88	6.87	.87	.61
Lower Extremity Function –Mobility*	20	56	95.65***		9.06	.77	.78
Upper Extremity Function -Fine Motor, ADL*	20	59	96.72***		8.34	.86	.94

* For these banks, a high score indicates better function; for all other banks a high score indicates worse function

** Time 1 (baseline) vs. Time 2 (7 days)

*** These two scales were not calibrated using IRT due to skewed distributions. Possible scores range from 0 (unable to do) -100 (without difficulty).

M_{GPT} – Mean General Population T-Score; M_{CT}- Mean Clinical T-Score

***4.8.6.3 Validity*.** Spearman rho correlations between the Neuro-QOL short forms and the pediatric disease measures are shown in Table 46 and between the Neuro-QOL short forms and the cross-disease instruments in Table 47.

Table 46. Correlations for Neuro-QOL short form T-scores with disease-specific measures

Neuro-QOL Short Form	PedsQL Core	PedsQL Emotional Functioning	PedsQL Physical Functioning	PedsQL Psychosocial Health	PedsQL School Functioning	PedsQL Social Functioning	MFS	MFS Cognitive Fatigue	MFS General Fatigue	MFS Sleep/Rest Fatigue
Depression	-.70***	-.66***	-.36**	-.68***	-.51***	-.49***	-.63***	-.59***	-.64***	-.47***
Anxiety	-.60***	-.51***	-0.19	-.55***	-.46***	-.37**	-.47***	-.44***	-.49***	-.39**
Stigma	-.50***	-.41**	-0.14	-.57***	-.42**	-.61***	-.34**	-.40**	-.36**	-0.14
Cognition	.53***	.41**	0.11	.53***	.52***	.35**	.57***	.66***	.53***	.30*
Lower Extremity Function - Mobility	-.46***	-.44***	-0.21	-.45***	-.28*	-.53***	-.40**	-.38**	-.45***	-0.21
Upper Extremity Function - Fine Motor, ADL	-.41**	-0.25	-0.18	-.38**	-.30*	-.46***	-.35**	-.39**	-.31*	-0.17
Fatigue	-.27*	-.30*	-0.06	-.32*	-.29*	-0.14	-.43***	-.46***	-.42***	-.26*
Pain	-.48***	-.48***	-0.25	-.46***	-.33*	-.28*	-.48***	-.43***	-.36**	-.45***
Social Relations – Interactions with Peers	.49***	.38**	0.18	.43***	0.22	.56***	.39**	.26*	.50***	.27*

*p < .05; **p < .01; ***p < .001

MFS = Multidimensional Fatigue Scale

Table 47. Correlations for Neuro-QOL short form T-scores with cross-disease measures

Neuro-QOL Short Form	Karnofsky Performance Scale	Symbol Digit Modalities # Correct	Symbol Search Raw Score	Digit Symbol Coding # Correct	PROMIS Physical Function T- Score	PROMIS Mental Health T-Score	Pain Scale (0-10)	EQ-5D Index Score	Global HRQL (0-4)
Depression	-.20	.08	-.10	.20	-.57***	-.71***	.23	-.32*	-.43***
Anxiety	-.16	.10	.01	.10	-.57***	-.60***	.19	-.33*	-.40**
Stigma	-.25	.01	-.15	.14	-.28*	-.34**	.01	-.37**	-.24
Cognition	.19	.16	.27*	.05	.42***	.52***	-.24	.46***	.29*
Lower Extremity (Mobility)	-.27*	.08	-.16	.17	-.36**	-.32*	.37**	-.42**	-.24
Upper Extremity (Fine Motor, ADL)	-.30*	-.17	-.45***	-.11	-.38**	-.30*	.38**	-.55***	-.14
Fatigue	-.09	.04	-.17	.12	-.36**	-.38**	.28*	-.49***	-.37**
Pain	-.25	-.13	-.08	.00	-.44***	-.35**	.57***	-.36**	-.40**
Social Relations – Interactions with Peers	.28*	.13	.12	.09	.45***	.34**	-.30*	.27*	.30*

*p < .05; **p < .01; ***p < .001

***4.8.6.4 Known Groups Validity*.** Patients with different seizure frequency (daily, weekly, monthly and yearly) scored significantly differently on Anxiety and Applied Cognition-General Concerns, with F=3.36, p=0.=25 and F=3.05, p=0.0358, respectively.

***4.8.6.5 Responsiveness*.** Similar to adult patients, we conducted responsiveness analyses on the Neuro-QOL banks using the Karnofsky Performance Status and the self-reported Global Rating of Change (GRC). Here we report the results from the GRC-based change. Beginning with the 7-level GRC (range: +3= very much better; 0 = about the same; -3 = very much worse), we collapsed the three "better" categories into one, and the three "worse" categories into one, leaving three categories ("better;" "about the same;" "worse"). These three categories were compared using one way analysis of variance followed by least significant difference testing of adjacent groups when the overall F statistic was significant. For each analysis, we required that at least 5 patients be represented in each of these three categories. If fewer than five patients were represented in a category, it was collapsed with the adjacent category and the two remaining groups were compared using a t-test. There were six GRC questions. Five of them queried patients specifically about change in Physical well-being, Cognitive well-being, Emotional well-being, Social/Family well-being, and Symptomatic Well-being (Disease-related Symptoms). The sixth GRC item asked about overall quality of life. The following indicates which of the 9 pediatric item bank change scores were compared across GRC categories:

Physical well-being: Physical Function (Upper and Lower extremity); Fatigue; Pain
Cognitive well-being: Applied Cognition - General Concerns
Emotional well-being: Depression; Anxiety; Stigma;
Social well-being: Social Relation- Interaction with peers; Stigma
Symptoms: Fatigue; Depression; Anxiety; Pain
Overall: ALL

This resulted in 23 planned comparisons for each wave two clinical validation sample (no adjustment made for multiple comparisons). Results for these responsiveness analyses are presented below. Only those that achieved statistical significance will be summarized. Of the 23 planned comparisons, two were statistically significant.

Emotional Well-being: Of the three planned comparisons, stigma was statistically significant ($F=3.24$, $p<0.05$). Post hoc comparisons showed that patients who reported a change (either better or worse) in Emotional Well-being at 6-month follow-up also reported higher stigma than did patients who reported no change in Emotional Well-being, effect size=0.53 and 0.78, respectively.

Social Well-being: Of the two planned comparisons, Stigma was found to be statistically significant ($t=2.02$; $p<.05$). Yet, the direction was unexpected. Patients who reported better Social Well-being at 6-months had more stigma than those who reported that their Social Well-Being was unchanged, with an effect size of 0.57.

***4.8.6.6 Conclusions*.** The current sample was generally high functioning. The 9 Neuro-QOL measures demonstrated high internal consistency. The Intraclass Correlation Coefficients (ICC) were acceptable, ranging from .44 (Stigma) to .94 (Upper Extremity Function- Fine motor, ADL). Convergent validity associations with generic and legacy measures were of the expected strength and direction. Responsiveness was not as good as we expected. It is hypothesized that this was due to the high functioning samples recruited in the testing with only a few patients reporting that they were getting worse at the 6-month follow-up.

4.8.7 RESULTS - MUSCULAR DYSTROPHIES

4.8.7.1 Sample characteristics. Patients (N=51) were primarily male (84.3%), white (58.8%), and non-Hispanic (62.7%) with average age=16.3 (SD=3.4; range=10.1 to 21.9). Seventy-seven percent were full time students, 2% were in school part time, and 4% were employed part-time. Of them, 5.9% (n=3) reported falling daily, 9.8% (n=5) weekly, 9.8% (n=5) monthly, 19.6% (n=10) rarely fall, yet 54.9% (n=28) were unable to ambulate without a wheelchair. One patient reported previous spine fracture, 11 (22%) limb fractures, and 17 (33.3%) received lower extremity or orthopedic surgeries before.

Mean T-Scores and standard deviations on the short forms are shown in Table 21. MD patients generally reported better functioning/ less symptom severity than the reference group norm with one exception. The exception was the Social Relations – Interactions with Peers Short Form, on which MD patients scored about 2.5 T-scores worse than the norm.

***4.8.7.2 Reliability*.** Internal consistency and 1 week test-retest reliability of the short forms is shown in Table 48. Cronbach's alphas range from .81 to .98 and ICCs from .61 to .97.

Table 48. Pediatric MD - Descriptive and reliability statistics for Neuro-QOL short form T-scores

Neuro-QOL Measures	N_{items}	$N_{persons}$	M_{GPT}	M_{CT}	*SD*	α	T-R** ICCs
Social Relations – Interactions with Peers*	8	50	47.42		10.15	.90	.87
Applied cognition: general concerns*	8	49		54.38	6.70	.81	.81
Depression	8	51	46.27		8.77	.92	.61
Anxiety	8	51	50.25		7.45	.85	.70
Stigma	8	51		49.29	7 .26	.92	.60
Fatigue	8	51		46.56	8.46	.81	.65
Pain	10	51		49.58	8.76	.92	.73
Lower Extremity (Mobility)* NOTE	20	22	54.02***		23.05	.90	.65
Upper Extremity (Fine Motor, ADL)*	20	51	53.63***		36.13	.98	.97

* For these banks, a high score indicates better function; for all other banks a high score indicates worse function
** Time 1 (baseline) vs. Time 2 (7-days)
*** These two scales were not calibrated using Item Response Theory models due to skewed distributions. Possible scores range from 0 -100
M_{GPT} – Mean General Population T-Score; M_{CT}- Mean Clinical T-Score

NOTE 28 patients (54.9%) reported using wheelchair only and had missing data on the Lower Extremity Function scale. When assigned "unable to do" for these patients on the Lower Extremity Function items, mean = 23.73.

4.8.7.3 Validity. Spearman rho correlations between the Neuro-QOL short forms and the pediatric disease measures are shown in Table 49 and between the Neuro-QOL short forms and the cross-disease instruments in Table 50.

Table 49. Correlations for Neuro-QOL short form T-scores with disease-specific measures

Neuro-QOL Short Form	PedsQL Core	PedsQL Emotional Functioning	PedsQL Physical Functioning	PedsQL Psychosocial Health	PedsQL School Functioning	PedsQL Social Functioning	Multidimensional Fatigue Scale (MFS)	MFS Cognitive Fatigue	MFS General Fatigue	MFS Sleep/Rest Fatigue
Depression	-.74***	-.74***	-.01	-.75***	-.59***	-.57***	-.58***	-.55***	-.59***	-.33*
Anxiety	-.70***	-.72***	-.13	-.72***	-.58***	-.46***	-.57***	-.48***	-.58***	-.40**
Stigma	-.73***	-.53***	.09	-.74***	-.52***	-.73***	-.48***	-.37**	-.51***	-.35*
Cognition	.60***	.46***	.11	.62***	.63***	.38**	.63***	.64***	.56***	.39**
Lower Extremity (Mobility)	-.20	-.12	.28	-.20	-.22	-.28	-.08	-.15	-.06	.12
Upper Extremity (Fine Motor, ADL)	-.04	-.19	-.31*	-.04	-.08	.08	.03	-.08	.01	.21
Fatigue	-.69***	-.51***	-.02	-.70***	-.63***	-.51***	-.65***	-.59***	-.62***	-.47***
Pain	-.73***	-.58***	.09	-.74***	-.57***	-.62***	-.74***	-.53***	-.65***	-.69***
Social Relations – Interactions with Peers	.41**	.40**	-.01	.42**	.41**	.32*	.36*	.38**	.37**	.13

*p < .05; **p < .01; ***p < .001

Table 50. Correlations for Neuro-QOL short form T-scores with cross-disease measures

Neuro-QOL Short Form	Karnofsky Performance Scale	Symbol Digit Modalities # Correct	Symbol Search Raw Score	Digit Symbol Coding # Correct	PROMIS Physical Function T- Score	PROMIS Mental Health T-Score	Pain Scale (0-10)	EQ-5D Index Score	Global HRQL (0-4)
Depression	-.05	-.40**	-.32*	-.35*	-.34*	-.70***	.27	-.20	-.40**
Anxiety	.04	-.19	-.22	-.30	-.35*	-.48***	.41**	-.20	-.28
Stigma	-.05	-.33*	-.41**	-.32*	-.42**	-.60***	.38**	-.23	-.25
Cognition	-.16	.29*	.21	.27	.37*	.41**	-.25	-.05	.26
Lower Extremity (Mobility)	-.62**	.01	-.22	-.18	-.28	-.32	-.05	-.37	-.10
Upper Extremity (Fine Motor, ADL)	-.82***	-.26	-.40**	-.45**	-.35*	-.29	-.20	-.72***	-.11
Fatigue	.32*	-.27	-.33*	-.26	-.40**	-.39**	.37**	.19	-.18
Pain	.23	-.34*	-.22	-.31*	-.51***	-.43**	.71***	-.26	-.15
Social Relations – Interactions with Peers	-.13	.47***	.27	.37*	.05	.49***	-.26	.15	.43**

*p < .05; **p < .01; ***p < .001

***4.8.7.4 Known Groups Validity*.** The global quality of life item "I am content with the quality of my life right now" (20.4% -Not at all or A little bit; 44.9% - Somewhat or Quite a bit; 34.7% - Very much) was used to evaluate known group differences of the pediatric Neuro-QOL measures. Depression, Anxiety, Applied Cognition-General Concerns and Social Relation-Interaction with Peers were statistically significant, F=7.32 (p=0.02), 3.51 (p=0.038), 3.59 (p=0.036) and 6.10 (p=0.005), respectively. Post-hoc comparisons showed that all significant comparisons were in the predicted direction, with effect size range from 0.75 to 1.58.

***4.8.7.5 Responsiveness*.** Same 23 planned comparisons as described in pediatric epilepsy were conducted. Results for these responsiveness analyses are presented below. Only those that achieved statistical significance will be summarized. Of the 23 planned comparisons, two were statistically significant.

Emotional Well-being: Of the three planned comparisons, Depression and Stigma were statistically significant, t= -2.29 (p=0.027) and t=-2.38 (p=0.022), respectively. Specifically, patients who reported "better" Emotional Well-being reported less depression and less stigma than those who reported it as remaining "the same". As less than 5 patients reported worsened Emotional Well-being at 6-month follow-up, these patients were grouped with "the same".

***4.8.7.6 Conclusions*.** The 9 Neuro-QOL measures demonstrated high internal consistency (alpha range from 0.81-0.98). The Intraclass Correlation Coefficients (ICC) were acceptable, ranging from .60 (Stigma) to .97 (Upper Extremity Function- Fine motor, ADL). Convergent validity with generic and legacy measures were of the expected strength and direction. Depression and Sigma were sensitive to change in Emotional Well-being change over time.

4.8.8 Overall Conclusions. These results summarize the procedures and initial findings from the Neuro-QOL clinical validation field testing. Overall, the Neuro-QOL short forms demonstrated excellent internal consistency across all diseases. Test-retest reliability was acceptable, but varied between disease groups. It was uniformly high for stroke, PD and MS, but a few short forms had lower than expected ICCs when used with ALS, adult and pediatric epilepsy, and muscular dystrophy patients. Validity of the Neuro-QOL short forms and scales was supported by 1). correlations with generic and disease-specific measures that were of the expected strength and direction; 2). Ability of the short forms to discriminate between patients grouped by disease severity level or other clinical factor. The majority of the Neuro-QOL Short Forms demonstrated adequate responsiveness to change over time.

5. CONCLUSIONS

Neurological clinical research will benefit from the availability of brief, reliable, valid and standardized questionnaires to measure health-related quality of life. The Neuro-QOL measurement system was funded for this purpose, and was created through extensive qualitative research with patients, caregivers and experts, using both classical and modern (e.g., IRT) test construction methodology. Its ongoing involvement and input from members of essential stakeholder groups (neurology and rehabilitation professionals, patients and caregivers) makes it a unique “patient-centered” tool that closely reflects recommended patient reported outcomes development guidelines from other agencies. The project has completed several tasks including: (1) identifying criteria for acceptance of neurology HRQL measures; (2) identifying target neurological diseases; (3) selecting HRQL domains and sub-domains for bank and scale creation; and (4) developing preliminary item pools and scales, (5) translating items into Spanish, (6) conducting large scale calibration testing with general population-based normative and clinical patient samples to gather psychometric data, calibrate items along a continuum, and create short forms and (7) testing calibrated short forms in a multisite clinical validation study.

Each validated Neuro-QOL short form comprises a set of items that have been carefully selected from the respective item banks to enhance estimation of a patient’s health status. Short forms (8-9 items each) are available for each domain and can be completed in less than 2 minutes by the typical patient. Researchers also may design their own short forms by selecting items from the item banks. These tools provide such an opportunity in a practical format for clinical research or practice. Over time and with accumulating publications, their use will be enhanced by increased interpretability with regard to the meaning of specific scores and score changes. The NINDS’s enthusiasm for these clinical research outcome measures is very high and as such, the potential impact of the Neuro-QOL measurement system on neurology clinical trials research is great. The Neuro-QOL system will continue to evolve, with additional banks and scales developed to capture HRQL issues not included in the current system that are important and useful for different patient populations.

6. REFERENCES

(1) Perez L, Huang J, Jansky L, Nowinski C, Victorson D, Peterman A 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(6):342-353.

(2) Eremenco SL, Cella D, Arnold BJ. A comprehensive method for the translation and cross-cultural validation of health status questionnaires. *Eval Health Prof.* 2005; 28(2):212-232.

(3) Reeve BB, Hays RD, Bjorner JB, Cook KF, Crane PK, Teresi JA et al. Psychometric Evaluation and Calibration of Health-Related Quality of Life Item Banks: Plans for the Patient-Reported Outcomes Measurement Information System (PROMIS). *Med Care.* 2007; 45(5 Suppl 1):S22-S31.

(4) Choi SW. Development of freeware for an iterative hybrid ordinal logistic regression/IRT DIF. 2009.
Ref Type: Unpublished Work

(5) Gershon R, Lai J-S, Bode R, Choi S, Moy C, Bleck T et al. Quality of Life Item Banks for Adults with Neurological Disorders: Item Development and Calibrations Based upon Clinical and General Population Testing. *To be submitted to Quality of Life Research.* 2010.

(6) Mahoney FI, Barthel DW. Functional evalutaion: The Barthel Index. *Maryland State Medical Journal.* 1965; 14:522-528.

(7) Granger CV, Albrecht GL, Hamilton BB. Outcome of comprehensive medical rehabilitation: Measurement by PULSES profile and the Barthel Index. *Arch Phys Med Rehabil.* 1979; 60(4):145-154.

(8) Shinar D, Gross CR, Bronstein KS, Licata-Gehr EE, Eden DT, Cabrera AR et al. Reliability of the activities of daily living scale and its use in telephone interview. *Arch Phys Med Rehabil.* 1987; 68(10):723-728.

(9) Lawton MP, Brody EM. Assessment of older people: Self-maintaining and instrumental activities of daily living. *Gerontologist.* 1969; 9(3):179-186.

(10) Schag CC, Heinrich RL, Ganz PA. Karnofsky performance status revisited: Reliability, validity, and guidelines. *Journal of Clincial Oncology.* 1984; 2(3):187-193.

(11) Lewandowski LJ. The Symbol Digit Modalities Test: a screening instrument for brain-damaged children. *Percept Mot Skills.* 1984; 59(2):615-618.

(12) Wechsler, D. (2006). *WISC-IV administration and scoring manual*. San Antonio, TX: The Psychological Corporation.

(13) Johnson JA, Coons SJ, Ergo A, Szava-Kovats G. Valuation of EuroQOL (EQ-5D) health states in an adult US sample. *Pharmacoeconomics.* 1998; 13(4):421-433.

(14) Rabin R, de CF. EQ-5D: a measure of health status from the EuroQol Group. *Ann Med.* 2001; 33(5):337-343.

(15) Hays RD, Bjorner J, Revicki DA, Spritzer K, Cella D. Development of Physical and Mental Health Summary Scores From the Patient Reported Outcomes Measurement Information System (PROMIS) Global Items. 2008.
Ref Type: Unpublished Work

(16) Cella D. *Manual of the Functional Assessment of Chronic Illness Therapy (FACIT Scales).* Evanston, IL: Center on Outcomes, Research and Education (CORE), Evanston Northwestern Healthcare and Northwestern University; 1997.

(17) Varni JW, Burwinkle TM, Katz ER, Meeske K, Dickinson P. The PedsQL in pediatric cancer: Reliability and validity of the Pediatric Quality of Life Inventory Generic Core Scales, Multidimensional Fatigue Scale, and Cancer Module. *Cancer.* 2002; 94(7):2090-2106.

(18) Varni JW, Seid M, Rode CA. The PedsQL: Measurement model for the pediatric quality of life inventory. *Med Care.* 1999; 37(2):126-139.

(19) Guyatt GH, Osoba D, Wu AW, Wyrwich KW, Norman GR. Methods to explain the clinical significance of health status measures. *Mayo Clin Proc.* 2002; 77(4):371-383.

(20) Guyatt GH, Norman GR, Juniper EF, Griffith LE. A critical look at transition ratings. *J Clin Epidemiol.* 2002; 55(9):900-908.

(21) Jaeschke R, Singer J, Guyatt GH. Measurement of health status. Ascertaining the minimal clinically important difference. *Controlled Clinical Trials.* 1989; 10(4):407-415.

(22) Guyatt GH, Berman LB, Townsend M, Pugsley SO, Chambers LW. A measure of quality of life for clinical trials in chronic lung disease. *Thorax.* 1987; 42(10):773-778.

(23) Osoba D, Brada M, Yung WKA, Prados M. Health-related quality of life in patients treated with temozolomide versus procarbazine for recurrent glioblastoma multiforme. *J Clin Oncol.* 2002; 18:1481-1491.

(24) Williams LS, Weinberger M, Harris LE, Clark DO, Biller J. Development of a stroke-specific quality of life scale. *Stroke.* 1999; 30(7):1362-1369.

(25) Kelly-Hayes M, Robertson JT, Broderick JP, Duncan PW, Hershey LA, Roth EJ et al. The American Heart Association Stroke Outcome Classification. *Stroke.* 1998; 29(6):1274-1280.

(26) Lai SM, Duncan PW. Evaluation of the American Heart Association Stroke Outcome Classification. *Stroke.* 1999; 30(9):1840-1843.

(27) Jenkinson C, Fitzpatrick R, Brennan C, Bromberg M, Swash M. Development and validation of a short measure of health status for individuals with amyotrophic lateral sclerosis/motor neurone disease: The ALSAQ-40. *Journal of Neurology.* 1999; 246 Suppl 3:III16-III21.

(28) Jenkinson C, Levvy G, Fitzpatrick R, Garratt A. The amyotrophic lateral sclerosis assessment questionnaire (ALSAQ-40): Tests of data quality, score reliability and response rate in a survey of patients. *Journal of the Neurological Sciences.* 2000; 180(1-2):94-100.

(29) Jenkinson C, Fitzpatrick R, Brennan C, Swash M. Evidence for the validity and reliability of the ALS assessment questionnaire: The ALSAQ-40. *Amyotrophic lateral sclerosis and other motor neuron disorders.* 1999; 1(1):33-40.

(30) Norquist JM, Fitzpatrick R, Jenkinson C. Health-related quality of life in amyotrophic lateral sclerosis: determining a meaningful deterioration. *Qual Life Res.* 2004; 13(8):1409-1414.

(31) Cedarbaum JM, Stambler N, Malta E, Fuller C, Hilt D, Thurmond B et al. The ALSFRS-R: a revised ALS functional rating scale that incorporates assessments of respiratory function. BDNF ALS Study Group (Phase III). *Journal of Neurological Sciences.* 1999; 169(1-2):13-21.

(32) Miller R, Bradley W, Cudkowicz M, Hubble J, Meininger V, Mitsumoto H et al. Phase II/III randomized trial of TCH346 in patients with ALS. *Neurology.* 2007; 69(8):776-784.

(33) Cutter GR, Baier ML, Rudick RA, Cookfair DL, Fischer JS, Petkau J et al. Development of a multiple sclerosis functional composite as a clinical trial outcome measure. *Brain.* 1999; 122(Pt 5):871-872.

(34) Kalkers N, Bergers L, de Groot V. Concurrent validity of the MS Functional Composite using MRI as a biological disease maker. *Neurology.* 2001; 56:215-219.

(35) Kalkers NF, Bergers L, de Groot V, Lazeron RH, van Walderveen MA, Uitdehaag BM et al. Concurrent validity of the MS Functional Composite using MRI as a biological disease marker. *Neurology.* 2001; 56(2):215-219.

(36) Fischer JS, Rudick RA, Cutter GR, Reingold SC. The Multiple Sclerosis Functional Composite Measure (MSFC): an integrated approach to MS clinical

outcome assessment. National MS Socieyt Clinical Outcomes Assessment Task Force. *Multiple Sclerosis.* 1999; 5(4):244-250.

(37) Benesova Y, Niedermayerova I. New approach to evaluation of clinical state in patients with multiple sclerosis. *Bratisl Lek Listy.* 2001; 102(8):370-371.

(38) Cohen JA, Cutter GR, Fischer JS, Goodman AD, Heidenreich FR, Jak AJ et al. Use of the multiple sclerosis functional composite as an outcome measure in a phase 3 clinical trial. *Arch Neurol.* 2001; 58(6):961-967.

(39) Hoogervorst EL, Kalkers NF, Uitdehaag BM, Polman CH. A study validating changes in the multiple sclerosis functional composite. *Arch Neurol.* 2002; 59(1):113-116.

(40) Rudick RA, Cutter G, Baier M, Fisher E, Dougherty D, Weinstock-Guttman B et al. Use of the Multiple Sclerosis Functional Composite to predict disability in relapsing MS. *Neurology.* 2001; 56(10):1324-1330.

(41) Rudick RA, Cutter G, Reingold S. The multiple sclerosis functional composite: a new clinical outcome measure for multiple sderosis trials. *Multiple Sclerosis.* 2002; 8(5):359-365.

(42) Fischer JS, LaRocca NG, Miller DM, Ritvo PG, Andrews H, Paty D. Recent developments in the assessment of quality of life in multiple sclerosis (MS). *Mult Scler.* 1999; 5(4):251-259.

(43) Cohen JA, Fischer JS, Bolibrush DM, Jak AJ, Kniker JE, Mertz LA et al. Intrarater and inerrater reliability of the MS functinoal composite outcome measure. *Neurology.* 2000; 54(4):802-806.

(44) Marrie RA, Goldman M. Validity of performance scales for disability assessment in multiple sclerosis. *Mult Scler.* 2007; 13(9):1176-1182.

(45) Nasreddine ZS, Phillips NA, Bedirian V, Charbonneau S, Whitehead V, Collin I et al. The Montreal Cognitive Assessment, MoCA: a brief screening tool for mild cognitive impairment. *J Am Geriatr Soc.* 2005; 53(4):695-699.

(46) de Boer AG, Wijker W, Speelman JD, De Haes JC. Quality of life in patients with Parkinson's disease: Development of a questionnaire. *Journal of Neurology, Neuroscience, and psychiatry.* 1996; 61(1):70-74.

(47) Peto V, Jenkinson C, Fitzpatrick R. PDQ-39: A review of the development, validation and application of a Parkinson's disease quality of life questionnaire and its associated measures. *Journal of Neurology.* 1998; 245 Suppl 1:S10-S14.

(48) Goetz CG, Fahn S, Martinez-Martin P, Poewe W, Sampaio C, Stebbins GT et al. Movement Disorder Society-sponsored revision of the Unified Parkinson's

Disease Rating Scale (MDS-UPDRS): Process, format, and clinimetric testing plan. *Mov Disord.* 2007; 22(1):41-47.

(49) Goetz CG, Poewe W, Rascol O, Sampaio C, Stebbins GT, Counsell C et al. Movement Disorder Society Task Force report on the Hoehn and Yahr staging scale: status and recommendations. *Mov Disord.* 2004; 19(9):1020-1028.

(50) Kroenke K, Spitzer RL, Williams JB. The PHQ-9: validity of a brief depression severity measure. *J Gen Intern Med.* 2001; 16(9):606-613.

(51) Devinsky O, Vickrey BG, Cramer J, Perrine K, Hermann B, Meador K et al. Development of the quality of life in epilepsy inventory. *Epilepsia.* 1995; 36(11):1089-1104.

(52) Cramer JA, Arrigo C, Van Hammee G, Bromfield EB. Comparison between the QOLIE-31 and derived QOLIE-10 in a clinical trial of levetiracetam. *Epilepsy Research.* 2000; 41(1):29-38.

(53) Smith DF, Baker GA, Dewey M, Jacoby A, Chadwick DW. Seizure frequency, patient-perceived seizure severity and the psychosocial consequences of intractable epilepsy. *Epilepsy Research.* 1991; 9(3):231-241.

(54) Rapp S, Shumaker S, Smith T, Gibson P, Berzon R, Hoffman R. Adaptation and evaluation of the Liverpool Seizure Severity Scale and Liverpool Quality of Life battery for American epilepsy patients. *Qual Life Res.* 1998; 7(4):353-363.

(55) Scott-Lennox J, Bryant-Comstock L, Lennox R, Baker GA. Reliability, validity and responsiveness of a revised scoring system for the Liverpool Seizure Severity Scale. *Epielpsy Research.* 2001; 44(1):53-63.

(56) Panelli RJ, Kilpatrick C, Moore SM, Matkovic Z, D'Souza WJ, O'Brien TJ. The Liverpool Adverse Events Profile: relation to AED use and mood. *Epilepsia.* 2007; 48(3):456-463.