



Published in final edited form as:

*Arch Phys Med Rehabil.* 2011 October ; 92(10 Suppl): S28–S36. doi:10.1016/j.apmr.2011.01.025.

## The Neurology Quality of Life Measurement Initiative

David Cella, Ph.D.<sup>1</sup>, Cindy Nowinski, MD, Ph.D.<sup>1</sup>, Amy Peterman, Ph.D.<sup>2</sup>, David Victorson, Ph.D.<sup>1</sup>, Deborah Miller, Ph.D.<sup>3</sup>, Jin-Shei Lai, Ph.D.<sup>1</sup>, and Claudia Moy, Ph.D.<sup>4</sup>

<sup>1</sup>Department of Medical Social Sciences, Northwestern University Feinberg School of Medicine, Chicago, IL

<sup>2</sup>Department of Psychology, University of North Carolina – Charlotte, Charlotte, NC

<sup>3</sup>Mellen Center for Multiple Sclerosis Treatment and Research, Cleveland Clinic Foundation, Cleveland, OH

<sup>4</sup>National Institute for Neurological Disorders and Stroke, National Institutes of Health, Bethesda, MD

### Abstract

**Objective**—The National Institute of Neurological Disorders and Stroke (NINDS) commissioned the Neurology Quality of Life (Neuro-QOL) project to develop a bilingual (English/Spanish), clinically relevant and psychometrically robust HRQL assessment tool. This paper describes the development and calibration of these banks and scales.

**Design**—Classical and modern test construction methodologies were used, including input from essential stakeholder groups.

**Setting**—An online patient panel testing service and eleven academic medical centers and clinics from across the United States and Puerto Rico that treat major neurological disorders.

**Participants**—Adult and pediatric patients representing different neurological disorders specified in this study, proxy respondents for select conditions (stroke and pediatric conditions), and English and Spanish speaking participants from the general population.

**Main Outcome Measures**—Multiple generic and condition specific measures used to provide construct validity evidence to new Neuro-QOL tool.

**Results**—Neuro-QOL has developed 14 generic item banks and 8 targeted scales to assess HRQL in five adult (stroke, multiple sclerosis, Parkinson's disease, epilepsy, and amyotrophic lateral sclerosis) and two pediatric conditions (epilepsy and muscular dystrophies).

**Conclusions**—The Neuro-QOL system will continue to evolve, with validation efforts in clinical populations, and new bank development in health domains not currently included. The potential for Neuro-QOL measures in rehabilitation research and clinical settings is discussed.

---

© 2011 The American Congress of Rehabilitation Medicine. Published by Elsevier Inc. All rights reserved.

Corresponding Author: David Cella, Ph.D. Department of Medical Social Sciences, Northwestern University Feinberg School of Medicine, 710 North Lake Shore Drive, Suite 729, Phone: (312) 503-1086, Fax: 312-503-9800, d-cella@northwestern.edu.

Part of the material in this manuscript was presented at the 135<sup>th</sup> annual meeting of the American Neurological Association (ANA), San Francisco, September 14, 2010.

I certify that no party having a direct interest in the results of the research supporting this article has or will confer a benefit or on any organization with which we are associated and I certify that all financial and material support for this research is clearly identified.

**Publisher's Disclaimer:** This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

## Keywords

Neurology; Clinical Research; Health-Related Quality of Life; Quality of Life; Patient Reported Outcomes

---

## Introduction

Neurologic disorders and their treatments can affect a wide array of physical, mental and social functioning, commonly referred to as health related quality of life (HRQL). Neuro-QOL is a new, standardized approach to measuring HRQL across common neurologic conditions. Since many neurologic conditions are chronic and incurable, treatment tends to focus on symptom management, limiting the extent of disability, and preventing disease progression. While there are some treatments that modify the course of these diseases, a major focus of management is rehabilitation. In short, treatment typically aims to improve the social, physical, and mental aspects 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. Multidimensional patient-reported outcome measures, such as 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 of disease modifying therapies and rehabilitation interventions, 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 item response theory (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, including rehabilitation services. This paper describes past accomplishments, current status and future plans for Neuro-QOL. All research activities reported in this paper received Institutional Review Board approval and all participants provided informed consent.

## Methods

### Identifying criteria for the acceptance of neurology HRQL measures

An early task was to gain understanding of what the neurology research community required in an HRQL measure in order to be interested in using it. This involved identifying objective criteria that should be met by the system. It also included an evaluation of investigator attitudes and beliefs that might need to be addressed in order to facilitate adoption. Since little is known about the factors influencing the use of HRQL measures in neurology, we modified an existing survey originally developed to examine use of HRQL data in oncology practice,<sup>1,2</sup> and used it to gather empirical information about the perspectives of neurologists and affiliated professionals regarding HRQL and HRQL instruments.

Drawing names from our consultant pool, a list of NINDS reviewers and grantees, and members of the American Academy of Neurology and the American Congress of Rehabilitation Medicine, we submitted a request for information to 719 neurology professionals. We received 103 responses (14%), with complete data available for item-level analysis on 89. The 89 responders reported a median age of 51 (33–89), were primarily male (70%), had practiced a median of 22 years, with the largest proportions coming from the professions of Neurology (47%) and Physiatry (15%). Sixty-seven (78%) experts saw only adult patients, 9% saw only pediatric patients, and 13% saw both. The vast majority (93%) had experience as an investigator in a clinical trial and reported having used HRQL measures (54%).

Sixty-six respondents provided qualitative data indicating HRQL measures should: 1) possess satisfactory psychometric properties (50% of all respondents); 2) be easy to administer and use (50%); 3) contain content reflecting the patient perspective and the diversity of symptoms and HRQL domains impacted by neurological disorders (27%); and 4) be clinically relevant and directly applicable to patient care (17%). Factor analysis of quantitative responses revealed two major perspectives (which we labeled Enthusiasm and Reluctance) that reflected positive or negative viewpoints toward HRQL. A median split on the enthusiasm and reluctance scales created four separate groups: high enthusiasm, low enthusiasm, high reluctance and low reluctance. Cross tabulations on these groups revealed four distinct patterns of respondents: *enthusiastic* (high enthusiasm/low reluctance; n= 25); *reluctant* (high reluctance/low enthusiasm; n=33); *uncommitted* (low reluctance/low enthusiasm; n=14) and *reluctantly enthusiastic* (high reluctance/high enthusiasm; n=17). Using a general linear model and Scheffe's post-hoc tests, we compared these four groups to determine the nature of any differences.

When compared to other groups those who were *enthusiastic* believed that HRQL can be objectively measured ( $p=.01$ ) and reported finding HRQL data more helpful in understanding their patients ( $p<.001$ ), and useful in changing their practice ( $p=.001$ ). Compared to other groups, *reluctant* respondents preferred focusing on clinical care over HRQL issues ( $p<.001$ ). The *uncommitted* and *reluctantly enthusiastic* groups were more likely to report willingness to use HRQL measures if they could be shown to be clinically relevant ( $p<.01$ ). Finally, *reluctantly enthusiastic* respondents were most likely to acknowledge that HRQL confirms clinical experience ( $p<.01$ ) and say that their use of HRQL measures would increase if they were easier to understand.

Taken together, these survey data suggested that incorporating those criteria identified from qualitative review, and in particular, ensuring that the Neuro-QOL system is clinically relevant and useful, easy to understand and to use will help support those who already feel generally positive toward HRQL measures and could help persuade those who are uncommitted or outright reluctant to use HRQL instruments.

## Selection of target conditions

A key element of the Neuro-QOL development strategy was the selection of the pediatric and adult conditions that would be used to test the assessment platform. We understood that this selection process needed to be inclusive and transparent, with significant input from the neurological research community. We intended to include neurological conditions that manifest across the normal human life span and had varying rates of morbidity and mortality. Results from each stage of this multi-step process are reported in Table 1.

The first step in the condition selection process involved an extensive literature review of neurological conditions in MEDLINE, PUBMED, Science Direct and Wiley Inter-science from 1996 to 2005 (when the review was completed). The search was conducted using combinations of key words including HRQL, neurological disorders, measurement issues and known disease-specific characteristics. This literature review was synthesized to identify conditions by their time of typical on-set, common health related quality of life concerns as well as disease-specific concerns and the likely impact of the condition on normal life span. Independent of this literature review, interviews were conducted with 44 experts in neurological disorders and/or health related quality of life to obtain their opinion about the 5 neurological conditions for which they felt it was most important to assess HRQL (see Table 1). They were not asked to specify whether they were nominating pediatric or adult conditions.

An expert consensus panel composed of 13 pediatric and adult neurology experts from across the country was convened in March, 2005, to establish and apply a set of criteria for selecting, per the NINDS contract, 5 adult and 2 pediatric conditions on which to build Neuro-QOL. After reviewing the results of the literature review and recommendations from the 44 individual expert reviews, members of this panel established criteria for selecting the 7 conditions which included: prevalence, individual impact, effective treatments, multiple domains affected, chronicity, and likelihood of HRQL change. Before the close of the consensus meeting, the panel nominated 5 adult and 2 pediatric conditions. An additional source of expert consultation was obtained when the results of the consensus meeting were presented to the American Academy of Neurology (AAN) for their comment. The recommended conditions from each step (interviews, consensus meeting and AAN) are presented in Table 1.

A final review of the recommended conditions was conducted with the NINDS staff and was reconciled with their historic grant portfolio. The final set of diseases, including their basis for inclusion, is presented in Table 1.

## Bank and Scale Development

**Identification of HRQL Domains and Sub-Domains**—The next step in our process was to determine which areas of HRQL to assess with the Neuro-QOL measures. We identified domains through multiple methods and data sources including a literature review, expert interviews, patient and caregiver focus groups and a keyword search.

**Literature Review:** First, we identified domains by completing an extensive Medline literature review of 24 major neurological conditions using key words such as health-related quality of life (HRQL), specific names of neurological disorders, measurement, as well as disease-specific characteristics, from 1996 to the present. This literature review summarized major neurological disorders and their impact upon HRQL, beginning with those typical to childhood onset followed by those most common in adults and advancing age. From this review, our initial list of domains included: emotional distress, perceived cognitive functioning, social functioning, physical functioning, fatigue, pain, communication/language

difficulty, positive psychological functioning, sexual functioning, bowel/bladder function, sleep disturbance and personality/behavioral changes.

**Expert Input:** We obtained expert input through two waves of expert interviews (n=44 and n=63 experts) and through the previously mentioned Request for Information (n=89) (see Table 2).

Experts were asked to identify domains or areas of HRQL that are affected by neurological disorders and their treatments. 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 people with neurological disorders. Experts were first asked to list all the domains they believed would be important to cover in an HRQL questionnaire that could be given to patients with neurological disorders (i.e., general and disease-specific). After that, they were asked to list domains that might be important in one of the disorders they named previously, but that weren't necessarily common to all disorders. During the individual interviews, experts provided greater depth and elaboration of content for given domains. For example, when the domain Physical Function was mentioned, experts may have elaborated further by mentioning activities of daily living, balance, fine motor skills, gait, hemiparesis, etc. Overall, these interviews confirmed domains that had been identified from the literature review and they also revealed the following new areas: behavior/personality change, driving, memory, attention, executive function, aggression/irritability, psychotic symptoms, meaning/spirituality and mastery/control.

**Patient and Caregiver Focus Groups:** We conducted eight focus groups with patients (total n=64) and three with caregivers (total n=19) to assess the impact of neurological conditions on HRQL domains. We began with broad questions, such as what do you think of when I say the phrase "quality of life" or "how has your life been affected by X condition?", allowing participants to freely list responses on their definition of quality of life as it relates to their health. We then progressed to questions regarding specific domains, such as physical function, emotional function, social aspects, and treatment effects that have been shown to be relevant in the literature. The previously mentioned focus groups with caregivers of Alzheimer's disease, stroke, and pediatric epilepsy patients were also conducted to gather important proxy perspectives from caregivers. Responses were qualitatively analyzed using NVivo software to determine the frequencies of each domain and sub-domain per diseases.<sup>3</sup>

**Key Word Search:** Because new domains arose from these different sources, we also conducted a comprehensive keyword literature search (from 1996 to 2005) using the OVID search engine with previous and newly identified domains and Neuro-QOL diseases to best estimate the number of published studies in a given area. We used these approximate totals to provide an overall quantification of how important certain domains were within different neurological conditions (see Table 3).

### Selection of HRQL Domains and Sub-Domains

After identifying the range of important domains and sub-domains, we selected the most important areas for item bank development. Working groups were formed for each of the seven Neuro-QOL conditions (stroke, adult epilepsy, ALS, Parkinson's disease, multiple sclerosis, muscular dystrophy, and pediatric epilepsy). Each group reviewed all data sources and extracted the most frequently-named and most relevant domains for item bank consideration.

Each source of data was analyzed using largely qualitative approaches. This process primarily entailed identifying and coding content derived from the previously described data



sources. 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 approach it was possible to understand how frequently physical function was mentioned in ALS, within the context of all other domains that were mentioned for ALS. This permitted a greater understanding of occurrence (and by association, importance) of certain domains either across all conditions or as a unique aspect of one disease. Frequent comparison to the literature and other sources of informant data were applied to enhance the data collection process.

Within each disease, domain percentages were calculated and recorded on a chart that was populated by information obtained from the various sources mentioned previously. For the expert input, to minimize experimenter demand and acquiescence biases, we included only the open-ended, spontaneously generated expert responses (vs. information experts suggested only after being asked to elaborate on a specific domain we provided them). If a domain was mentioned across all five data sources (e.g., literature review, 3 types of expert input, focus groups, key word search), it received a score of “5”; if it was mentioned across four data sources, it received a score of “4”, and so on. These 0–5 counts were then compared across diseases. If a domain was counted as  $\geq 3$  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  $\geq 2$  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 4 for generic and targeted domains. After reviewing the findings of this comprehensive identification and selection process, the generic domains that were chosen for item bank development were: Physical, Social, Emotional and Cognitive Function.

Next, we identified domain co-chairs from the Neuro-QOL Executive Committee and co-investigator panel. Each co-chair team was assigned a domain from the four generic domains previously selected and one pair was assigned to oversee the targeted domains. Each dyad was charged with reviewing the aforementioned data sources and extracting the most relevant subdomains for item bank consideration. Due to funding restraints, a decision was made by the Executive Committee to develop and test up to three targeted banks, and develop but not test others, thus providing future investigators with item pools that could be subsequently advanced. Frequent checks back with NINDS to keep the project anchored to the original scope afforded us useful feedback regarding relevance, vis-à-vis the original purpose of the project, which was to create psychometrically robust patient reported outcomes of HRQL that could be used by neurology clinical trials researchers. Data were analyzed using the approaches described below.

Using data from expert interview domain elaborations, we calculated the percentage of times a particular code was applied within a domain. This helped us estimate which codes might carry additional importance for a particular domain within a disease based on how often they were discussed among experts. 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. If an issue was present across a majority of diseases, it was labeled as generic. The following generic sub-domains were selected for item bank development in adults: Physical (Self-care/Upper Extremity, Mobility/Ambulation), Social (Role Participation, Role Satisfaction), Emotion (Depression, Anxiety, Positive Psychological Function), Cognitive (Perceived, Applied). In pediatrics, the following generic sub-domains were selected for item bank development: Physical (Self-care/Upper Extremity, Mobility/Ambulation), Social, Emotion (Emotional Health, Stigma).

Based on feedback from experts, as well as considering the complexity of issues surrounding these conditions, we decided to develop and field test one (1) targeted scale per condition, and also develop (but not field test) additional targeted scales as indicated by the unique circumstances of each condition. To determine which scales would be field tested, we summarized and examined data from our data sources in which domain elaboration were available. Using these data we made preliminary decisions regarding which targeted scales should be developed, and for which disease(s). This led to the identification of a select number of candidate domains, which were presented to disease specific experts involved in the Neuro-QOL study. 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, focus group frequency counts, 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 only (see Table 5).

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 5, ALS, MD, MS and PD all appear to have greater numbers of bowel and bladder issues that were coded, compared to adult/pediatric epilepsy, and stroke.

### Identifying and selecting existing items

For each of the domains and sub-domains selected as a critical part of the HRQL universe for neurological disorders, large pools of relevant items were identified from a variety of sources. An extensive, iterative process took place with the goals of obtaining comprehensive coverage of each content area, then selecting a "best set" of items for field testing.

Candidate items for the generic item banks and targeted scales were identified from our existing item banking projects and affiliated studies, Rasch analysis of several large external datasets, and additional generic and disease-specific questionnaires that have been used in neurological conditions. Permission from outside principal investigators and primary scale authors was obtained for the latter two activities. These data were evaluated by examining the content and dimensionality of the constituent items in these preliminary banks.

From these various data sources, a centralized Neuro-QOL Item Library was created. Over 3,000 items were entered into this Library according to elements such as item order, context, time frame, item stem and response options. An extensive "binning" and "winnowing" process was then undertaken. This iterative, multi-step process involved at least three domain experts. Two of these independent raters worked collaboratively to assign items to "bins" according to primary domain. After this, a third rater reconciled any discrepancies. As the number of items (many redundant) was quite large, all items were reviewed to determine if they should proceed through detailed item review/revision/testing. Items were then grouped together according to each domain's hierarchy of sub-domains, factors and facets. Once all items were assigned to a domain, content experts "winnowed" (i.e., systematically removed) items from item pools. Items were removed for a variety of reasons, including semantic redundancy, availability of a superior alternative, inconsistency with domain definition, wrong domain assignment, vague or confusing language, gender inappropriateness, narrow applicability, and likelihood of problems in cultural/linguistic

translation. Remaining items were then reviewed by two Neuro-QOL investigators 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; and ease of translation.

### **Qualitative item review and cognitive interviews**

The comprehensive item pool for each HRQL domain was then subjected to a qualitative item review (QIR) process. 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.

**Expert item review (EIR)**—Before cognitive interviews were conducted with patients, every item in the comprehensive pool was reviewed by at least three experts for clarity, precision, acceptability to respondents, adaptation to computerized testing, format of responses, preferred response options and similarity of timeframe. Two Neuro-QOL domain experts then evaluated that information and made decisions about the need for review or modification of individual item. Expert collaborators: a) signed off on items that appeared to need no further revision; and b) suggested revisions to items that still needed improvement. The final item pools were approved after review by members of the Neuro-QOL Executive Committee.

**Cognitive interviews**—After identifying approximately the 50 best items per generic item bank or disease-specific scale, cognitive interviews were conducted by telephone with 63 adult and pediatric patients with Neuro-QOL conditions, as well as four 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.

### **Final steps to creation of field test-ready item banks and scales**

Because the items would be translated into Spanish, it was important to consider problems that might arise during that translation. Accordingly, translation science experts provided feedback about the ease of translating all items and potential item response categories (e.g., “not at all” to “very much”): this information was used to modify items, when possible; to remove items that appeared to be particularly problematic for translation; and to choose the final response categories for the various types of items (e.g., frequency, severity).

Each domain working group carefully reviewed all the input from neurology experts, patients and translation scientists and made appropriate changes. The proposed final, field-test ready item banks and scales were reviewed by all the working group and domain chairs. The Neuro-QOL Executive Committee gave final approval prior to the first field test.



## Spanish language version

From the outset, one of this project's aims was to make all of the item banks/scales readily available for use in the Spanish-speaking population. Input was obtained from native Spanish speaking patients with neurological disorders in all the previous steps for which patient input was solicited. A rigorous forward-backward translation process<sup>4</sup> was undertaken to translate the field test-ready item banks and scales described above. Following this extensive work to obtain a high quality linguistic translation, the items were cognitively debriefed with 30 adults and 30 children. Each subject was asked to first answer a subset of the translated items independently. Next, a Spanish speaking interviewer asked the subject about the meaning of specific words within the item stem, 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. On the basis of the cognitive interviews, some revisions were made to the original translations.

## Results

### Item calibration testing and short form construction

**Testing Sample and Associated Domains**—To obtain reliability and validity data on scales, and item calibrations on banks, we conducted two waves of initial testing. Table 6 details the testing by domain and provides initial psychometric data.

The first wave (Wave Ia) was a test of targeted scales. By their nature, these scales are specific in their content to issues germane to clinical populations. Therefore, the targeted scales were first tested in their relevant clinical populations. Respondents in this sample were recruited by an Internet-based opt-in panel, YouGovPolimetrix ([www.polimetrix.com](http://www.polimetrix.com), also see [www.pollingpoint.com](http://www.pollingpoint.com)), a polling firm based in Palo Alto, CA. A total of 511 adults and 50 children were recruited in Wave Ia. For adults, the average age was 56.2 (SD=12.8) years, 53% were male, and 95% were white. Of the 511 adults, 209 had a diagnosis of stroke, 183 epilepsy, 84 MS, 50 PD, and 18 ALS (a person could have more than one diagnosis). For children, the average age was 14.4 (SD=1.9), 51% were male, 92% were white, and 97% attended school. Fifty of the children had a diagnosis of epilepsy and 9 had MD.

The remaining domains were calibrated in Wave Ib testing using the US general population. This sample was recruited by another internet panel company, [www.greenfield.com](http://www.greenfield.com). In consideration of respondent burden, subjects were asked to complete only 2–3 item banks (i.e., no more than 100 items) and therefore, sample sizes for each bank varied (shown in table 6).

**Analysis**—Data from each domain were analyzed separately. In addition to basic statistics such as alpha and item-total correlations (see Table 6), we evaluated dimensionality of items included in each bank using factor analytic techniques. Various factor analytic techniques (criteria are detailed in Reeve et al, 2007<sup>5</sup> and Lai et al, 2006<sup>6</sup>) were used, including exploratory factor analysis (EFA), one-factor analysis (CFA) and bi-factor analysis. Depending on the nature of the domain, more than one technique might be used. For example, in pediatric emotional health, we evaluated the dimensionality of items from both the psychometric perspective as well as by taking the clinical perspective into account. From the psychometric perspective, one item bank including all items from depression, anxiety, worry and anger was acceptable. This conclusion was based on satisfactory one-factor CFA results (comparative fit index, CFI = 0.92) and high inter-factor correlations (range: 0.839–0.943) found when a three-factor CFA was conducted (CFI = 0.94). However, different intervention strategies have been used for treating depression and anxiety and therefore,

these two concepts traditionally have been evaluated separately. Therefore, we decided to build two separate item banks for depression and anxiety (CFI=0.97 for each of the banks analyzed separately). Items that satisfied unidimensionality requirements were retained and further evaluated by using  $S-\chi^2$  and  $S-G^2$  fit indices as developed by Orlando and Thissen.<sup>7</sup> Finally, item parameters were estimated using the Graded Response Model<sup>8</sup> as implemented in MULTILOG.

We applied the above approaches to all item banks/scales except four pediatric disease specific scales administered in Wave Ia, where only 59 children were recruited. Due to the sample size limitation, analysis focused only on descriptive statistics. Rasch analysis,<sup>9</sup> which required a smaller sample size, was also used for exploratory purposes with the understanding that item parameters are likely to be changed when a different sample is tested.

We then created short-forms for the item banks, but not disease specific scales, to be used for Wave II clinical validation. There are many methods to construct short-forms and more than one short-form can be created. For this study, one short-form was created for each domain, and items included in each short-form were selected by using multiple indices and determined in a consensus meeting. The indices included item precision (i.e., information function produced by IRT analysis), locations on the measurement continuum to ensure representativeness across the measurement continuum, IRT fit indices, frequency of being selected in CAT simulation, frequency counts, and clinical importance. Due to the skewed distributions found for mobility/ambulation and fine motor/upper extremity function for both adults and children, the study group decided to select items for the Wave II validation by consulting experts with reference to the analysis results. Short-form item length is indicated in Table 6.

## Evaluation of Neuro-QOL in Clinical Populations – Wave II

We are currently evaluating the validity, reliability and responsiveness of Neuro-QOL short forms and disease specific scales with people suffering from the target diseases. We are enrolling 500 adults across five clinical conditions with 100 proxies matched to the Stroke sample, and 100 children across two clinical conditions, with another 100 proxies matched to the pediatric sample. Within each disease, males and females will be recruited proportionally to the gender breakdown within that disease.

Physician ratings, administration of concurrent measures and/or chart review will be conducted at baseline and as part of the 180-day follow up sample. All patient groups will also receive disease-specific measures to evaluate validity and responsiveness.

We anticipate that baseline assessments will be complete by January, 2010, with follow-up assessments finished by July, 2010. Results will be analyzed to evaluate reliability, validity and sensitivity to change with the final instruments ready for public dissemination in September 2010. Table 7 shows the item banks, short forms (SF) and disease specific scales (DSS), along with the approximate number of items in each, that we expect to be available at that time. However, analysis results may lead to some modifications. CAT algorithms for each item bank will also be available, although CATs will not yet be implemented.

## Discussion

### Connections to Other Projects and Implications for Rehabilitation Medicine

Throughout Neuro-QOL, we have made every effort to build upon and forge connections to already existing HRQL assessment efforts. In particular, Neuro-QOL has strong links to two well-developed and accepted measurement systems; the NIH Patient Reported Outcome

Measurement Information System (PROMIS; [www.nihpromis.org](http://www.nihpromis.org)) and the Activity Measure for Post-Acute Care (AM-PAC)<sup>10</sup>. Once Neuro-QOL domains were selected, it became apparent that considerable conceptual overlap existed between Neuro-QOL and both of these efforts. PROMIS and AM-PAC items were extensively reviewed by teams of domain specific clinical content experts with experience in neurological disorders, quality of life and other chronic illnesses. Many of these items, with permission, were incorporated into Neuro-QOL's generic item pools. While some items needed re-writing, ranging from minor modifications to a complete overhaul; a sufficient number of items remained for future linking efforts. (See article within this issue describing linking between Neuro-QOL and AM-PAC).

## Study Limitations

Neuro-QOL begins, but does not complete, the process of developing and validating a comprehensive, efficient measurement system for patient-reported outcomes in neurology clinical research. We were limited in the diseases that could be addressed and the domains that could be measured. Further research can continue to provide validation of these initial item banks and scales, and extensions into other disease and QOL domains.

## Conclusions

Efforts have been made to link the Neuro-QOL tool to the larger field of rehabilitation medicine, as for example, the AM-PAC project noted above. There are also several government funded extensions of the Neuro-QOL measurement tool, most notably in the areas of spinal cord injury (SCI) and traumatic brain injury (TBI). NINDS and National Institute on Disability and Rehabilitation Research NIDRR funded studies are currently underway to expand Neuro-QOL into SCI. Wherever possible, common items from generic domains (e.g., emotional health) link both efforts for future cross walking purposes, while new SCI-specific content covers important disease targeted areas, such as physical-medical complications like respiratory difficulties or autonomic dysreflexia. NIDRR and Department of Veterans Affairs VA funded efforts are also on-going to accomplish similar global goals; however in TBI, tools are being developed and tested both with those injured from the general population as well as returning wounded warriors from Iraq and Afghanistan. Neuro-QOL study team members have been involved on all of these extensions to insure conceptual and methodological equivalence. These expansions into the field of rehabilitation medicine have considerable potential for improving health outcomes measurement in that field. Similarly, standardized HRQL evaluations such as Neuro-QOL can influence patient care and healthcare policy, by improving assessment of patient-reported outcomes and disease burden in neurological diseases, increasing consistency in measurement across rehabilitation and neurology research, and offering a common metric that provides a common language to express burdens of disease and benefits of treatment, as they are experienced by the patient.

## Acknowledgments

This study was supported by contract # HHSN265200423601C from the National Institute of Neurological Disorders and Stroke. Reprints will not be available from the authors.

## Abbreviations

<b>HRQL</b>	health related quality of life
<b>NINDS</b>	National Institute of Neurological Disorders and Stroke

<b>IRT</b>	Item Response Theory
<b>CAT</b>	Computer Adaptive Test
<b>AAN</b>	American Academy of Neurology
<b>EFA</b>	Exploratory Factor Analysis
<b>CFA</b>	Confirmatory Factor Analysis
<b>SF</b>	Short Form
<b>DSS</b>	Disease Specific Scale
<b>PROMIS</b>	Patient Reported Outcomes Measurement Information System
<b>AM-PAC</b>	Activity Measure for Post Acute Care
<b>EIR</b>	Expert Item Review
<b>QIR</b>	Qualitative Item Review

## References

1. Taylor KM, Macdonald KG, Bezjak A, Ng P, DePetrillo AD. Physicians' perspective on quality of life: An exploratory study of oncologists. *Qual Life Res.* Feb; 1996 5(1):5–14. [PubMed: 8901361]
2. Bezjak A, Taylor KM, Ng P, MacDonald K, DePetrillo AD. Quality-of-life information and clinical practice: The oncologist's perspective. *Cancer Prev Control.* Oct; 1998 2(5):230–235. [PubMed: 10093637]
3. Perez L, Huang J, Jansky L, et al. Using focus groups to inform the Neuro-QOL measurement tool: exploring patient-centered, health-related quality of life concepts across neurological conditions. *J Neurosci Nurs.* Dec; 2007 39(6):342–353. [PubMed: 18186419]
4. 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. [PubMed: 15851774]
5. Reeve BB, Hays RD, Bjorner JB, 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.* May; 2007 45(5 Suppl 1):S22–S31. [PubMed: 17443115]
6. Lai JS, Crane PK, Cella D. Factor analysis techniques for assessing sufficient unidimensionality of cancer related fatigue. *Qual Life Res.* Sep; 2006 15(7):1179–1190. [PubMed: 17001438]
7. Orlando M, Thissen D. Further examination of the performance of  $S-X^2$ , an item fit index for dichotomous item response theory models. *Applied Psychological Measurement.* 2003; 27:289–298.
8. Samejima, F.; van der Liden, WJ.; Hambleton, R. *Handbook of modern item response theory.* New York, New York: Springer; 1996. The graded response model; p. 85-100.
9. Wright, BD.; Masters, GN. *Rating scale analysis: Rasch measurement.* Chicago: MESA Press; 1985.
10. Haley SM, Coster WJ, Andres PL, et al. Activity outcome measurement for postacute care. *Med Care.* Jan; 2004 42(1 Suppl):I49–I61. [PubMed: 14707755]

**TABLE 1**  
Expert Nominations and Rationale for Selection of Neurological Conditions as Research Priorities

Diseases Nominated	Nominating Groups			Final Selected Conditions	Rationale
	Individual Expert Interviews	Consensus Group	American Academy of Neurology Practice Committee		
Stroke	x	x	x	x	Support from literature and nominated across all groups
Multiple Sclerosis	x	x	x	x	
Parkinson's disease	x	x	x	x	
Amnrotrophic lateral sclerosis	x			x	Support from literature and recommended by NINDS to include a neuromuscular condition with prominent HRQL impact
Epilepsy (Adults)	x		x	x	Support from literature and majority of nominating groups; Provides opportunity to study one condition across the life span
Epilepsy (Pediatrics)	x	x	x	x	
Muscular Dystrophies		x		x	Support from literature, Consensus Panel and NINDS Input
<hr/>					
Alzheimer's Disease and dementias	x	x	x		
Migraine Headache (Adults)	x	x			
Traumatic Brain Injury (Adults)	x				
Traumatic Brain Injury (Pediatrics)	x				
Migraine Headache (Pediatrics)	x				
Cerebral Palsy			x		

NOTE: Conditions listed above dotted line were selected as Neuro-QOL conditions



**TABLE 2**

## Expert Background and Experience

	Interview I (n=44)	Interview II (n=63)	Online Request for Information (n=89)
Years in Practice (median)	20	21	22
Male	70%	70%	70%
Profession			
Neurology	57%	43%	47%
Physiatry	14%	18%	15%
Health/Rehab Psychology	7%	9%	8%
Neuropsychology	7%	7%	8%
Nursing	4%	2%	1%
Other	11%	21%	21%
Adult patients only	70%	78%	78%
Pediatric patients only	16%	8%	9%
Both	14%	14%	13%
Investigator in a clinical trial	89%	89%	93%
Use HRQL scales in research	73%	56%	54%
Use HRQL scales in practice	75%	29%	29%

**TABLE 3**

Example of Keyword Literature Search

	Published studies	ALS	Multiple Sclerosis	Pediatric Epilepsy	Adult Epilepsy	Parkinson's Disease	Stroke	Muscular Dystrophy
		2,851	9,709	8,972	6,001	11,591	20,352	776
<u>PHYSICAL</u>								
Fine/Gross motor skill	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
Activities of Daily Living	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

TABLE 4

Domains and Importance Scores Across Diseases

	Adult Epilepsy	MS	Stroke	PD	ALS	Pediatric Epilepsy	MD	Generic or 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 Symptoms	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  $\geq 3$  in 50% of diseases; Targeted =  $\geq 2$  in less than 5 diseases; MS = Multiple Sclerosis; PD = Parkinson's Disease; ALS = Amyotrophic lateral sclerosis; MD = Muscular Dystrophy

**TABLE 5**

Targeted Scales for Development and Field Testing

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

TABLE 6

## First Wave Testing of Generic Item Banks and Targeted Scales

Bank or Scale	Domain	Adult (A) or Pediatrics (P)	Sample N	# of items tested	# of items retained	Alpha	Item-total corr	# of items included in short-form
<i>Wave 1a Testing</i>								
Scale	Sleep Disturbance	A	511	20	20	.92	.39-.70	20
Scale	Personality and Behavior Changes	A	511	20	18	.95	.49-.84	18
Scale	Stigma	A; P	511; 59	26; 20	24; 18	.97; .98	.53-.83 .71-.93	24; 18
Scale	Fatigue/Weakness	A; P	511; 59	20; 13	19; 13	.98; .97	.53-.89 .58-.90	19; 13
Scale	Cognition	P	59	20	19	.97	.57-.87	19
Scale	Pain	P	59	10	10	.97	.86-.94	10
<i>Wave 1b Testing</i>								
Bank	Depression	A; P	513; 513	37; 19	30; 18	.98; .97	.64-.90 .52-.88	8; 8
Bank	Anxiety and Fear (A)/Worry (P)	A; P	513; 513	28; 19	28; 19	.97; .97	.56-.87 .62-.86	8; 8
Bank	Positive Psychological Function	A	513	27	27	.98	.60-.91	9
Bank	Perceived Cognitive Function	A	513	48	46	.98	.57-.85	20
Bank	Applied Cognitive Function	A	513	42	42	.97	.54-.78	20
Bank	Mobility and Ambulation	A; P	549; 505	37; 39	37; 39	.97; .98	.41-.79 .50-.87	20; 20
Bank	Fine Motor/Upper Extremity Function	A; P	549; 505	44; 40	44; 40	.97; .98	.45-.76 .40-.87	20; 20
Bank	Role Performance	A	549	49	45	.99	.66-.91	8
Bank	Role Satisfaction	A	549	51	45	.99	.53-.89	8
Bank	Social Function (interaction w/peers; w/adults)	P	513	38	24	.95; .92	.45-.84 .45-.84	8; 8



**TABLE 7**

## Neuro-QOL Banks, Short Forms and Disease-Specific Scales

<b>Domain</b>	<b># of Items in Bank (Adult/Pediatric)</b>	<b># of Adult Items</b>	<b># of Pediatric Items</b>	<b>Form*</b>
Depression	31/18	8	8	SF
Anxiety/Fear	28/19	8	8	SF
Stigma	24/18	8	8–10	SF
Positive Psychological Function	27	9	--	SF
Perceived Cognitive Function	47	8–10	--	SF
Applied Cognitive Function	42	8–10	--	SF
Mobility and Ambulation	37/39	8–10	8–10	SF
Fine Motor/Upper Extremity Function	44/40	8–10	8–10	SF
Role Performance	49	8	--	SF
Role Satisfaction	51	8	8	SF
Social Function	25	--	8	SF
Cognition	--	--	18	DSS
Fatigue/Weakness	--	19	13	DSS
Sleep Disturbance	--	20	--	DSS
Personality and Behavior Changes	--	18	--	DSS
Pain	--	--	10	DSS