Linking the Activity Measure for Post-acute Care and the Quality of Life Outcomes in Neurological Disorders

Stephen M. Haley, Ph.D.[*], **Pengsheng Ni, MD, MPH**[*], **Jin-shei Lai, Ph.D.**[‡], **Feng Tian, M.S.**[*], **Wendy J. Coster, Ph.D.**[†], **Alan M. Jette, Ph.D.**[*], **Donald Straub, MS. PT**[§], and **David Cella, Ph.D.**[‡]

[*]Health and Disability Research Institute, Boston University School of Public Health, Boston, MA

[†]Department of Occupational Therapy, Sargent College of Health and Rehabilitation Sciences, Boston University, Boston, MA

[‡]Northwestern University Feinberg School of Medicine, Chicago

[§]Department of Physical Therapy, University of Illinois at Chicago

Abstract

Objective—To use item response theory (IRT) methods to link physical functioning items in the Activity Measure for Post-acute Care (AM-PAC) and the Quality of Life Outcomes in Neurological Disorders (Neuro-QOL)

Design—Secondary data analysis of the physical functioning items of AM-PAC and Neuro-QOL. We used a non-equivalent group design with 36 core items common to both instruments. We used a test characteristic curve transformation method to for linking AM-PAC and Neuro-QOL scores. Linking was conducted so that both raw scores and scaled AM-PAC and Neuro-QOL scores (converted-logit scores with mean = 50 and SD = 10) could be compared.

Setting—AM-PAC items were administered to rehabilitation patients in post-acute care settings. Neuro-QOL items were administered to a community sample of adults via the Internet.

Participants—The AM-PAC sample consisted of 1,041 post acute care patients; the Neuro-QOL sample was 549 community-dwelling adults.

Interventions—Not applicable.

Main Outcome Measures—25 Mobility items and 11 ADL items common to both instruments were included in the analysis.

Corresponding Author: Alan M Jette, PT, PhD, Director, Health & Disability Research Institute, Acting Chairman, Department of Health Policy & Management, BU School of Public Health, 715 Albany St., T5W, Boston, MA 02118, Phone: 617 638-1985, ajette@bu.edu.

Results—Neuro-QOL items were linked to the AM-PAC scale using the Generalized Partial Credit Model. Mobility and ADL subscale scores from the two instruments were calibrated to the AM-PAC metric.

Conclusions—An IRT-based linking method placed AM-PAC and NeuroQOL Mobility and ADL scores on a common metric. This linking allowed estimation of AM-PAC Mobility and ADL subscale scores based on Neuro-QOL Mobility and ADL subscale scores, and vice versa. The accuracy of these results should be validated in a future sample in which participants respond to both instruments.

Keywords

Introduction

An anticipated feature of contemporary patient reported outcome (PROs) instruments for clinical populations has been the ability to link items of old and new instruments by calibrating scores to a common metric.[1, 2] Placing items from different PRO instruments on a common metric and developing a linkage between their scores allows us to identify scores on different measures that have comparable meaning. These procedures are fundamental to creating a "network" of assessments that can be compared with one another in future clinical research.[3]

Traditional raw-score or classical test theory (CTT) approaches have been used to develop cross-walk tables for instruments used in rehabilitation medicine, including comparing the Functional Independence Measure (FIM) and Minimal Data Set (MDS) items and scores[4] and the FIM and MDS-Post Acute Care (PAC) scores.[5, 6] Both studies yielded only marginal results, with important discrepancies found in certain item-to-item matches. Objections to using a CTT approach include potential error from developing item-to-item matches based on expert panels and test-dependency with the individual samples selected for the linking study.[7]

More recently, the linking of instruments in rehabilitation medicine has been accomplished using Item Response Theory (IRT) models. Assuming the PRO constructs of different instruments are measuring the same dimension, IRT methods develop correspondence tables and figures to describe the association of equivalent scores for different instruments. In order to overcome test dependency, both instruments are placed onto a common metric. Under IRT, linking can be accomplished by transforming the item parameter estimates from one test to the metric of a second test. For example, McHorney and colleagues[8] successfully used an IRT approach to link three physical functioning item on a common continuum and Velozo and colleagues[7] linked the FIM and MDS using Rasch methods and conducted a validation study.[9]

A number of current projects are attempting to build scoring links by selecting common items that will be administered in both instruments. For example, a number of core items have been developed and are being tested in two separate item bank development projects, one for children[10] with spinal cord injury (SCI), and one for adults with SCI.[11] The goal is to link the two instruments so that scores from the pediatric spinal cord injury instrument can be related to scores in the adult SCI instrument for comparability in long-term follow-up. Other projects that will provide score linking opportunities include work by Tulsky and colleagues, who are co-calibrating Quality of Life Outcomes in Neurological Disorders (NeuroQOL) items[12] with items specifically developed for persons with spinal cord injury.

In this study, we chose to link two instruments because of their common content domains and items.

IRT linking studies include both a sampling strategy and a linking strategy.[13] Sampling procedures consist of common subjects, common items, or some combination of common items and subjects. Linking procedures include either putting all items on a single scale, (common calibration) or using calibrations from common items to create a link between the two instruments.[2] In the current project, we had available data from two instruments with common items, but different samples, yielding a nonequivalent group sampling design. The use of common items for subsequent linking was part of the initial design and development of Neuro-QOL. Neuro-QOL incorporated a number of items from the Activity Measure for Post-acute Care (AM-PAC) within its Lower Extremity Function (Mobility) and Upper Extremity Function (Fine motor, ADL) item pools. Although the initial calibration work with Neuro-QOL used a community-dwelling adult database and the AM-PAC used a sample of post-acute care patients, having a set of core items that were administered in both samples allowed for linking the instruments. Our linking design used linking coefficients from common items to create corresponding scoring tables. This linking design was chosen because of adequate sample size availability[14] and the desire to maintain the original calibrations of the AM-PAC, which had been previously calibrated. The purposes of this article are to demonstrate the use of a nonequivalent sampling design using linking coefficients from common items to develop score conversions for the physical functioning domains of the AM-PAC and Neuro-QOL.

METHODS

Sample and Procedures

Our methods included a non-equivalent groups design with common items administered to both community-dwelling and clinical samples. In the development of the scoring links, we defined the common items and used a test characteristic curve transformation method to develop linking coefficients for the conversion of scores to the same scale. Test characteristic curves, are mathematical equations that estimate total score on the same measure—conditional on participants' levels of the trait being measured. Pairs of scores from two different instruments are equivalent if they correspond to the same level of the trait being measured.

Using a non-equivalent sampling approach, we drew upon data from two samples. For the AM-PAC, data were collected from 1,041 participants who were 18 years and older from six regional rehabilitation networks. Exclusion criteria included: (1) non-English- speaking patients and (2) patients who were in a coma, debilitated, or agitated to a degree that precluded participation in active rehabilitation. AM-PAC data were collected by semi-structured interviews, either by the participant's clinician, or by a trained data collector. Each interview (approximately 45–60 min) was fully scripted, with standard instructions, practice questions, and an answer card to help subjects communicate response choices. A subject's ability to respond to self-report questions was assessed by the treating clinician or assigned data collector who determined if the participant could: (1) understand the interview questions; (2) sustain attention for an hour; and (3) reliably respond to questions. Full details of the data collection procedures have been previously reported.[15]

The Neuro-QOL sample (N=549) consisted of adults without neurological conditions who were part of an Internet-based opt-in panel, YouGovPolimetrix (www.polimetrix.com), also see www.pollingpoint.com), a polling firm based in Palo Alto, CA. YouGovPolimetrix operates PollingPoint.com, a centralized portal that allows interested individuals to provide their views about public policy and other current issues. The respondents for a typical

YouGovPolimetrix Internet survey are selected from the PollingPoint panel, a panel of over one million respondents who have provided YouGovPolimetrix with their names, street addresses, email addresses, and other information, and who regularly participate in online surveys. Panelists were recruited by a variety of methods including e-random digit dialing, invitations via web newsletters, and Internet poll-based recruitment where panelists have opted to participate in a survey advertised on the World Wide Web. Panel members receive modest compensation (less than $10 value) when they participate.

Physical Functioning Measures

AM-PAC—The AM-PAC measures functional activity in adults across all post-acute care settings. Early content and analytic work with the AM-PAC established three major activity content domains upon which the AM-PAC dimensions are based (1) Basic Mobility, (2) Daily Activities, and (3) Life Skills.[16] Only analyses of the Basic Mobility and Daily Activity items are reported in this article. A number of versions of the AM-PAC instrument have been developed, with the current version based on a sample of 1,041 PAC patients. The AM-PAC uses a 4-point difficulty rating scale (no difficulty, a little difficulty, a lot of difficulty, cannot/unable to do). Both AM-PAC short-forms and CAT formats have been developed. We have conducted a series of simulated validation tests of the AM-PAC CAT software on numerous PAC samples, including patients with stroke,[17] complex medical conditions,[18] and prospective work on patients with orthopedic conditions.[19, 20] We have found the AM-PAC to be responsive to short (3-mos)[21] and long-term (12-month) changes in post acute care patients.[22] A staging system to help characterize the meaning of scores has also been developed.[23] We used the original set of community-dwelling calibrations for the AM-PAC, and have removed only walking aid device items (rare in the Neuro-QOL normative sample) from the analysis. In order to develop consistency with parallel Neuro-QOL domains, we will refer to the AM-PAC Basic Mobility Scale as *Mobility*, and the AM-PAC Daily Activity scale as *ADL.*

NeuroQOL—Neuro-QOL is a 5-year, multi-site NINDS funded project that aims to develop a clinically relevant and psychometrically robust health-related quality of life (HRQL) assessment tool that will be responsive to the needs of researchers in a variety of neurological disorders for both adults and children and settings and facilitate comparisons of data across clinical trials in different diseases. Physical function item banks were among those developed under Neuro-QOL.

The Neuro-QOL physical function item banks consist of two banks: Upper Extremity Function (Fine motor, ADL) and Lower Extremity Function (Mobility). Contents of each bank were solicited based on results of focus groups and individual interviews with experts. We (JSL, DS) reviewed all potential items, mainly from PROMIS; www.nihpromis.org and AM-PAC, removed items that tapped similar content, and arranged items based on their hypothesized difficulty levels on the physical function continuum. New items were written if suspected gaps were identified on the continuum or if important content was missing. Resulting item pool was reviewed by two experts, and then reviewed by the NeuroQOL team. Finally, items were reviewed by patients with one of the following neurological conditions: stroke, multiple sclerosis, ALS, Parkinson's disease and epilepsy. Specifically, each item was reviewed by at least 5 patients to ensure its readability. As a result, the final Lower Extremity Function (Mobility) item bank consists of 37 items while Upper Extremity Function (Mobility) item bank consists of 44 items measured on a 5-point difficulty rating scale (either "Without any Difficulty, With a Little Difficulty, With Some Difficulty, With Much Difficulty, Unable to Do" or "No Difficulty, A Little Difficulty, Some Difficulty, A Lot of Difficulty, Can't Do"). We will refer to the Neuro-QOL Lower Extremity Function

(Mobility) items as *Mobility* and the Upper Extremity Function (Fine motor, ADL) items as *ADL.*

Data Analysis

Most IRT models applied to PROs measures assume that the item responses constitute a unidimensional dataset. [24, 25] The unidimensionality of the two AM-PAC domains of Mobility and ADL has been previously documented.[16, 26] therefore, no dimensionality analyses were conducted for the AM-PAC data. Item parameter estimates for the AM-PAC items were obtained in previous work.[15] The dimensionality of responses to the Neuro-QOL items, (Mobility and ADL scales, separately) was evaluated using categorical exploratory factor analysis (EFA). Because the items were polytomous, we used an unweighted least squares (ULS) estimator based on a polychoric correlation matrix. The eigenvalues, scree plots and percentage of variance explained were used to assess unidimensionality. Confirmatory factor analysis (CFA) also was used to test unidimensionality. CFA model fit was assessed by multiple fit indexes, including Comparative Fit index (CFI), Tucker–Lewis Index (TLI) and Root Mean Square Error Approximation (RMSEA). CFI and TLI compare the model to a baseline null model; possible values range from 0 to 1; 0.95 or higher suggests acceptable fit. RMSEA assesses misfit per degree of freedom; values less than 0.08 means acceptable fit.[24]

A number of IRT models are appropriate for use in calibrating responses to polytomous items (i.e., items with more than two response categories.[27] To obtain item parameters for Neuro-QOL, we fit the Generalized Partial Credit Model (GPCM) to the Mobility and ADL scales separately. We modeled the Neuro-QOL item responses using the GPCM because this model was used in the original AM-PAC data. We calibrated items with the marginal maximum likelihood estimation using PARSCALE.[28] Item fit was tested by the likelihood ratio test[28] and Stone's G-statistics test;[29] p values less than 0.05 indicated item misfit.

The common 44 items included among both the AM-PAC and Neuro-QOL items (27 Mobility and 15 ADL) were analyzed for differential item function (DIF). DIF occurs when factors other than the ability of the person influence responses. For example, at similar functional levels, men report less pain than women. For the current study we examined DIF with respect to the two samples; that is, we evaluated whether any items functioned differently in the AM-PAC (post-acute patients) and in the Neuro-QOL (normative) samples.

There are two kinds of DIF. One is uniform DIF in which item response probabilities differ consistently across all trait levels between groups (AM-PAC/Neuro-QOL samples), that is, there is no interaction between the group membership and the trait level. The second is nonuniform DIF in which item response probabilities varies between groups along the trait levels, that is, there is interaction between the group membership and the trait level. DIF was assessed using ordinal logistic regression. The dependent variable was the responses to an item. Three independent variables include the trait level measured with the total raw score of the common items, the group membership and an interaction term between the trait level and the group membership. The analytic strategy is to successively add the trait level, the group membership and the interaction term into the model and the procedure was repeated for each item. A significant group membership effect indicates the presence of uniform DIF and a significant interaction term indicates non-uniform DIF. Model comparisons were based on the likelihood ratio test. The effect size of the DIF was classified based on the R-square change between models.[30]

We set the following criteria for the DIF analysis. An R-square change less than 0.035 was judged to indicate unimportant DIF; R-square change between 0.035–0.07 to indicate

moderate DIF; and R-square change greater than 0.07 to indicate substantial DIF and such items would be removed from the linking process.[30] An important purpose of these DIF analyses is to identify a subset of the common items that do not exhibit substantial DIF. It is these DIF-free items that serve at the common link between the items of the two scales.

We used the Stocking-Lord method, a test characteristic curve approach to linking.[31] The Stocking-Lord strategy is to take the parameters of the items common to both measures and apply a transformation to them such that the test characteristics curve of the "new measure" (Neuro-QOL) is as similar as possible to the test characteristic curve of the "old measure" (AM-PAC). To accomplish this, linking coefficients (denoted as *A*, *B*) are estimated. These transformation constants are chosen to minimize the weighted sum of squared distances between the two test characteristic curves (based on the items common between the two measures). Once the best values of *A* and *B* were found, they were used to transform the remaining Neuro-QOL items (the ones not in common with AM-PAC) onto the AM-PAC scale. The standard errors of linking coefficients were calculated based on the delta method.[32] Once the Neuro-QOL items were on the same metric as the AM-PAC, we constructed item maps that compared item difficulty parameters with the distribution of the domain in the samples. We then associated Neuro-QOL Mobility and ADL domain scores with their comparable AM-PAC scores.

Results

Sample Characteristics

AM-PAC sample—The AM-PAC calibration data were collected from 1,041 participants who were 18 years and older (mean = 63.6 yrs; SD=16.6; range = 18–100 yrs.). Most subjects were women (56.8%), white (86.5%), and unmarried (61.9%). All patients were actively receiving post acute care (PAC) services at the time of AM-PAC assessment and were from one of four settings: inpatient rehabilitation (n = 420), skilled nursing facility (SNF) or transitional care units (TCU) (n = 138), home care (n = 246), and outpatient (n = 237). Patients were categorized into three major impairment groups based on standard rehabilitation impairment codes: (1) 32.7% neurologic (e.g., stroke multiple sclerosis, Parkinson's disease, brain injury, spinal cord injury, neuropathy); (2) 35.9% musculoskeletal (e.g., fractures, joint replacements, orthopedic surgery, joint, or muscular pain); and (3) 31.4% medically complex (e.g., debility resulting from illness, cardiopulmonary conditions, or postsurgical recovery). Full details of sample characteristics have been previously reported.[15]

Neuro-QOL sample—The average age of the 549 participants in this sample was 52.9 (SD=14.7). About half (49%) were male; 53% were married; 23% had a high school education or less; 33% were employed full-time and 32% were retired. Most (91% were White; 6% were Hispanic; 6% were Black/African American; and the remaining 4% were Asian, American Indian/Alaskan Native, or Native Hawaiian/Pacific Islander. Most participants (74%) reported none of the neurological conditions listed. The most frequently reported condition was migraines (13%). The most frequently reported comorbidities were: hypertension (43%) depression (29%), anxiety (22%), diabetes (20%) and sleep disorder (18%). See Table 1 for a summary of the demographics of the AM-PAC and Neuro-QOL samples.

Unidimensionality

Exploratory factor analysis (EFA) of the 37 Mobility items and 44 ADL Neuro-QOL items revealed two factors that explained 59% of the variance for Mobility and 79% of the item variance for ADL, indicating that Mobility and ADL can be considered two separate

domains. The Mobility domain model fit was generally acceptable CFI=0.982, TLI=0.996, RMSEA=0.10). The NeuroQOL ADL scale was better fitting. (CFI=0.96, TLI=0.991, RMSEA=0.09).

Results of likelihood ratio test showed that the Neuro-QOL Mobility items (the ratio of chi-square to df =1.14) and ADL (the ratio of chi-square to df =1.35) fit the generalized partial credit model. Only two Mobility items ('running 45 minutes' and 'running up and down an incline') and one ADL item ('tie your shoelaces') showed misfit in both the likelihood ratio test and in Stone's G statistic.

DIF of Common Items

We started with 42 common items: 27 Mobility and 15 ADL items. Three items in Mobility were classified as moderate DIF and one item ('getting into and out of a truck with your wheelchair') as level large DIF. In Mobility, we also removed items ('moving from sitting at the side of the bed to lying down on your back') due to the high residual correlation with other items. Four items in ADL had moderate DIF and 4 items ('taking off a pullover shirt', 'chopping or slicing vegetables', 'shaving your neck and face safely and thoroughly with an electric razor' and 'holding a screw and screwing it in tight with a manual screwdriver') had level large DIF. Our final set of common items included 25 Mobility and 11 ADL items.

Linking

We used the Stocking-Lord method[33] to link items that minimized the difference in test characteristic curves in common AM-PAC and Neuro-QOL Mobility items and between common AM-PAC and Neuro-QOL ADL items. The differences in linking coefficients were statistical significant, indicating that there were differences in test characteristic curves between the common items of the two instruments. See Table 2.

We present in Figures 1 and 2 the item maps for the Mobility and ADL items placed on a common (AM-PAC) scale. Item maps present the content range of the item banks for which item difficulty scores and person ability scores are plotted. For each item, the estimated midpoint ability level for each of the core items and additional AM-PAC and Neuro-QOL items are located on the item map. In the Mobility domain, the common items were located in the middle of the continuum. AM-PAC had many more items requiring less ability than Neuro-QOL, The remaining Neuro-QOL items tended to be at the higher end of the Mobility continuum, indicating they reflected greater capabilities. In the ADL item map, the common items were again clustered in the mid-range of the scale. The remaining Neuro-QOL items were in the moderate ability range, while the AM-PAC items were spread more evenly across the continuum, with some items at the higher ability range.

Figure 3 displays the linear transformation for the transformed-logit scaled scores (AM-PAC scale Mean = 50; SD = 10) for both the Mobility and ADL domains of AM-PAC and Neuro-QOL. For example, if a person had a transformed score of 40 on the Mobility domain of Neuro-QOL, the corresponding score on AM-PAC would be 57. A transformed score of 30 on Neuro-QOL converts to a score of 41 in ADL. We have created a similar conversion illustration in Figure 4 to convert raw scores (sum of the rating scale values of each item) between instruments. For example, a raw Mobility score of 120 on Neuro-QOL is comparable to a 253 raw score on the AM-PAC; a Neuro-QOL ADL raw score of 60 is comparable to a raw score of 86 on the AM-PAC. Other conversion illustrations could be developed including raw score conversions either to the AM-PAC or Neuro-QOL scaled scores.

Discussion

Although linking methods have been used for years in education,[26] they have been less frequently applied in health status assessment. Previously, the proliferation of PROs that differed in breadth and depth of measurement made it impossible to draw comparisons across measures that were used in different studies. Despite years of advances in classical outcome measurement, data from two studies or populations assessing the same outcomes could not be compared when different set of items were used. This made it impossible to compare treatment effects across different studies. One important advantage of contemporary outcome measurement is the ability to compare test scores from different measures of the same PROs by linking different PRO tests. This linking procedure allows one to relate scores from different measures on a common metric, thus facilitating cross study comparisons.

This study demonstrated that both the Mobility and ADL domains of the AM-PAC and Neuro-QOL measures could be linked and score conversion tables could be developed. The results revealed that by linking these two instruments, scores generated from the Neuro-QOL Upper and Lower Extremity Function item banks can be compared and contrasted with those generated by the AM-PAC Basic Mobility and Daily Activity item banks.

Linking was accomplished both with raw scores as well as scores transformed to a T-score distribution. Since scores provided in contemporary PROs are often expressed as T-scores, with a mean of 50 and standard deviation of 10; the provision of linking tables for the AM-PAC and Neuro-QOL instruments enhances the comparability of both of these PROs. The original person scores generated from IRT software are in log-odd (logits) units. Since logit scores are not widely used by clinicians and researchers, to ease interpretation, raw logit scores are transformed to the T-score to enable consistent interpretations by having a mean score of 50 and standard deviation of 10.[34] The data in Figure 1 provide an illustration of how Neuro-QOL and AM-PAC were converted using the T-score metric (often termed scaled scores). Since the conversion of scaled scores between Neuro-QOL and AM-PAC is linear, the conversion formulae are also provided in Figure 1. However for those working with raw scores, the data in Figure 2 illustrate raw score conversion between these two PROs. Since raw scores are not on a linear scale, their conversion is also not linear.

Study Limitations

There are several limitations in these analyses that should be noted. First when only a small number of items are available for the anchor set in linking, there is a risk of not being able to obtain precise estimates of ability. As Kolen[35] argues, the common items between instruments being linked should comprise 20–30% of all the items in the domain, although we know of no empirical data to support this. In these analyses the common items comprised only 9–16% of the full item banks due, in part, to the need to remove some common items that had significant DIF between the community NeuroQOL and clinical AM-PAC samples. Furthermore, to the extent that the common items are bunched together on the ability scale, estimates of ability are also likely to be less precise. Ideally, the common items across the two PRO instruments should be spread out across the full continuum of content. Additionally, a fundamental requirement of linking scores is that the two instruments are measuring the same content dimension. As seen in Figures 1 and 2, the common items tended to cluster in the mid-range of functional abilities. This may have had some effect on the accuracy of the linking coefficients.

Conversion information can be developed in both directions- either starting with a raw or scaled Neuro-QOL score and converting to an AM-PAC score, or starting with a raw or scaled AM-PAC score and converting to a Neuro-QOL score. Since the original linking

coefficients were developed in the direction of converting Neuro-QOL to AM-PAC scores, the score conversion in this direction is likely to be more accurate. We did examine a few cases to see how well the backward conversion (AM-PAC to Neuro-QOL) worked. Using the example from Figure 3, the ADL conversion from a Neuro-QOL score of 30 to AM-PAC was 42.3. If one started from an AM-PAC score of 42.3, the Neuro-QOL conversion is 30.1. Similarly for Mobility, conversion from a Neuro-QOL score of 58 to AM-PAC was 40.4. If one started from an AM-PAC score of 40.4, the Neuro-QOL conversion is 40.0. These differences are trivial and suggest that the conversion is likely to be robust in both directions. This assumption will need to be tested in future validation work. The Neuro-QOL data were obtained in a community-based sample and were intended to serve as the reference group to other clinical populations. Therefore, future work should evaluate using Neuro-QOL as the anchor test from which other tests are compared. As the number of items changes or new content is added to either instrument, these conversion results will need to be updated

Conclusions

We have demonstrated the ability to link the scores between the NeuroQOL and AM-PAC Mobility item banks and between the NeuroQOL and AM-PAC ADL item banks. Common items between the two instruments permitted us to develop linking coefficients that provided score conversion information. In the future, conversion tables can be provided across all forms of an instrument (such as short-forms or computer adaptive testing formats). This work provides an example of linking instruments with common items and non-equivalent samples and demonstrates what might be possible as newly developed instruments retain content from existing measures.

Acknowledgments

List of Abbreviations

TCU	Transitional Care Units
IRT	Item Response Theory
AM-PAC	Activity Measure for Post-acute Care
Neuro-QOL	Quality of Life Outcomes in Neurological Disorders
ADL	Activities of Daily Living
FIM	Functional Independence Measure
MDS	Minimal Data Set
PRO	Patient Reported Outcome
CTT	Classical Test Theory
SCI	Spinal Cord Injury
CAT	Computer Adaptive Test
HRQL	Health-Related Quality of Life
PROMIS	Patient Reported Outcome Measurement Information System

ALS	Amyotrophic Lateral Sclerosis
EFA	Exploratory Factor Analysis
ULS	Unweighted Least Squares
CFA	Confirmatory Factor Analysis
CFI	Comparative Fit index
TLI	Tucker–Lewis Index
RMSEA	Root Mean Square Error Approximation
GPCM	Generalized Partial Credit Model
DIF	Differential Item Function
PAC	Post-Acute Care
SNF	Skilled Nursing Facility

References

1. Hambleton, RK. Applications of Item Response Theory to Improve Health Outcomes Assessment: Developing Item Banks, Linking Instruments, and Computer-Adaptive Testing. In: Lipscomb, J.; Gotay, CC.; Snyder, C., editors. Outcomes Assessment in Cancer. Cambridge, UK: Cambridge University Press; 2005. p. 445-64.

2. Dorans NJ. Linking scores from multiple health outcome instruments. Qual Life Res. 2007; 16 (Suppl 1):85–94. [PubMed: 17286198]

3. Cella D, Gershon R, Lai J, Choi S. The future of outcomes measurement: item banking, tailored short-forms, and computerized adaptive assessment. Quality of Life Research. 2007; 16(1):133–41. [PubMed: 17401637]

4. Williams BC, Li Y, Fries BE, Warren RL. Predicting patient scores between the Functional Independence Measure and the Minimum Data Set: development and performance of a FIM-MDS "crosswalk". Archives of Physical Medicine and Rehabilitation. 1997; 78(1):48–54. [PubMed: 9014957]

5. Buchanan JL, Andres PL, Haley SM, Paddock SM, Zaslavsky AM. An assessment tool translation study. Health Care Financing Review. 2003; 24(3):45–60. [PubMed: 12894634]

6. Buchanan JL, Andres PL, Haley SM, Paddock SM, Zaslavsky AM. Evaluating the planned substitution of the Minimum Data Set-Post Acute Care for use in the rehabilitation hospital prospective payment system. Med Care. 2004; 42:155–63. [PubMed: 14734953]

7. Velozo CA, Byers KL, Wang YC, Joseph BR. Translating measures across the continuum of care: Using Rasch analysis to create a crosswalk between the Functional Independence Measure and the Minimum Data Set. J Rehabil Res Dev. 2007; 44(3):467–78. [PubMed: 18247243]

8. McHorney CA. Use of item response theory to link three modules of functional status items from the Asset and Health Dynamics among the Oldest Old Study. Archives of Physical Medicine and Rehabilitation. 2002; 83(3):383–94. [PubMed: 11887121]

9. Wang Y, KLB, CAV. Validation of FIM-MDS crosswalk conversion algorithm. J Rehabil Res Dev. 2008; 45(7):1065–76. [PubMed: 19165695]

10. Calhoun C, Haley S, Riley A, Vogel L, McDonald C, Mulcahey M. Development of items designed to evaluate activity performance and participation in children and adolescents with spinal cord injury. International Journal of Pediatrics. submitted.

11. Slavin M, Kisala P, Jette ADST. Developing a contemporary outcome measure for spinal cord injury research. Spinal Cord. in press.

12. Cella D, Young S, Rothrock N, Gershon R, Cook K, Reeve B, et al. The Patient-Reported Outcomes Measurement Information System (PROMIS): progress of an NIH Roadmap

Cooperative Group during its first two years. Medical Care. 2007; 45(5 Suppl):S3–S11. [PubMed: 17443116]

13. Chen W, Revicki D, Lai J, Cook K, Amtmann D. Linking pain items from two studies onto a common scale using item response theory. Journal of Pain Symptom & Management. 2009; 38(4): 615–28.

14. Cook K, Taylor P, Dodd B, Teal C, McHorney C. Evidence-based practice for equating health status items: sample size and IRT model. Journal of Applied Measurement. 2007; 8(2):175–89. [PubMed: 17440260]

15. Haley SM, Ni P, Hambleton RK, Slavin MD, Jette AM. Computer adaptive testing improves accuracy and precision of scores over random item selection in a physical functioning item bank. J Clin Epidemiol. 2006; 59:1174–82. [PubMed: 17027428]

16. Haley SM, Coster WJ, Andres PL, Ludlow LH, Ni PS, Bond TLY, et al. Activity outcome measurement for post-acute care. Med Care. 2004; 42(Suppl 1):I-49–I-61. [PubMed: 14707755]

17. Andres PL, Black-Schaffer RM, Ni PS, Haley SM. Computer adaptive testing: a strategy for monitoring stroke rehabilitation across settings. Topics in Stroke Rehabilitation. 2004; 11(2):33–9. [PubMed: 15118965]

18. Siebens H, Andres P, Ni P, Coster W, Haley S. Measuring physical function in patients with complex medical and postsurgical conditions: a computer adaptive approach. American Journal of Physical Medicine & Rehabilitation. 2005; 84(10):741–8. [PubMed: 16205429]

19. Haley S, Ni P, Jette A, Tao W, Moed R, Meyers D, et al. Replenishing a computerized adaptive test of patient-reported daily activity functioning. Quality of Life Research. 2009; 18(4):461–71. [PubMed: 19288222]

20. Jette A, Haley S, Tao W, Ni P, Moed R, Meyers D, et al. Prospective evaluation of the AM-PAC-CAT in outpatient rehabilitation settings. Physical Therapy. 2007; (87):385–98. [PubMed: 17311888]

21. Haley S, Siebens H, Coster W, Tao W, Black-Schaffer R, Gandek B, et al. Computerized adaptive testing for follow-up after discharge from inpatient rehabilitation: I. Activity outcomes. Archives of Physical Medicine & Rehabilitation. 2006; 87:1033–42. [PubMed: 16876547]

22. Coster WJ, Haley SM, Jette AM. Measuring patient-reported outcomes after discharge from inpatient rehabilitation settings. Journal of Rehabilitation Medicine. 2006:237–42. [PubMed: 16801206]

23. Tao W, Haley S, Coster W, Ni P, Jette A. An exploratory analysis of functional staging using an item response theory approach. Arch Phys Med Rehabil. 2008; 89:1046–53. [PubMed: 18503798]

24. 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):S22–31. [PubMed: 17443115]

25. Thissen D, Reeve BB, Bjorner JB, Chang CH. Methodological issues for building item banks and computerized adaptive scales. Qual Life Res. 2007; 16 (Suppl 1):109–19. [PubMed: 17294284]

26. Coster WJ, Haley SM, Andres PL, Ludlow LH, Bond TLY, Ni PS. Refining the conceptual basis for rehabilitation outcome measurement: personal care and instrumental activities domain. Med Care. 2004; 42(Suppl 1):I-62–I-72. [PubMed: 14707756]

27. van der Linden, W.; Hambleton, R. Handbook of Modern Item Response Theory. Berlin: Springer; 1997.

28. PARSCALE [Computer Program]. Lincolnwood, IL: Scientific Software International Inc; 2003.

29. Stone CA. Empirical power and Type I error rates for an IRT fit statistic that considers the precision of ability estimates. Educational and Psychological Measurement. 2003; 63:566–83.

30. Jodoin MG, Gierl MJ. Evaluating type I error and power rates using an effect size measure with the logistic regression procedure for DIF detection. Applied Measurement in Education. 2001; 4(14): 329–49.

31. Kolen, MJ.; Brennan, RL. Test equating, scaling, and linking: methods and practices. 2. New York: Springer; 2004.

32. Ni P, Haley S, Jette A. Standard errors of item response theory (IRT) equating/linking in polytomous IRT models Applied Phsychological Measurement submitted.

33. Stocking ML, Lord FM. Developing a common metric in item response theory. Applied Psychological Measurement. 1983; 7:201–10.

34. Ware JE Jr, Gandek B, Sinclair SJ, Bjorner B. Item response theory in computer adaptive testing: implications for outcomes measurement in rehabilitation. Rehabilitation Psychology. 2005; 50(1): 71–8.

35. Kolen MJ. Linking assessments effectively: purpose and design. Educational Measurement: Issues and Practice. 2001; 20(1):5–s9.
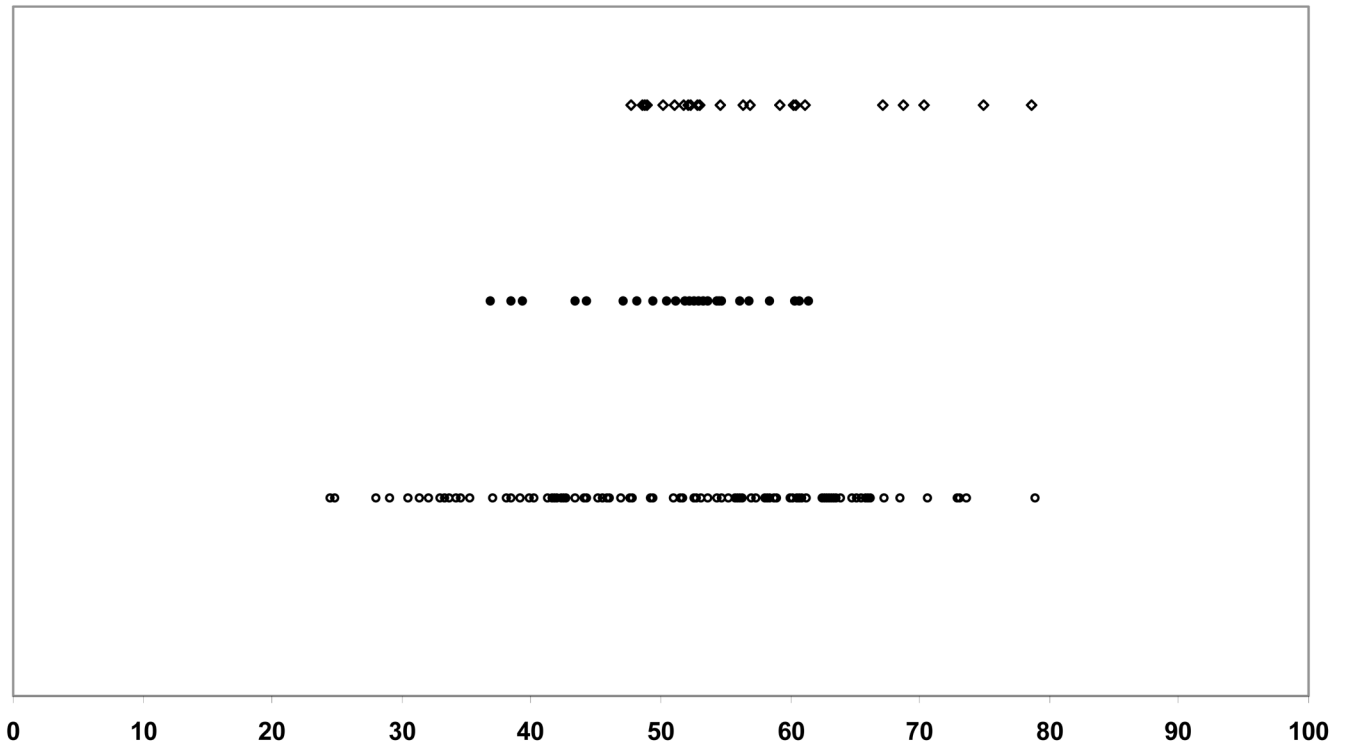
Figure 1.
Item Map of Mobility items (Common items and remaining AM-PAC and NeuroQOL items) on AM-PAC scale
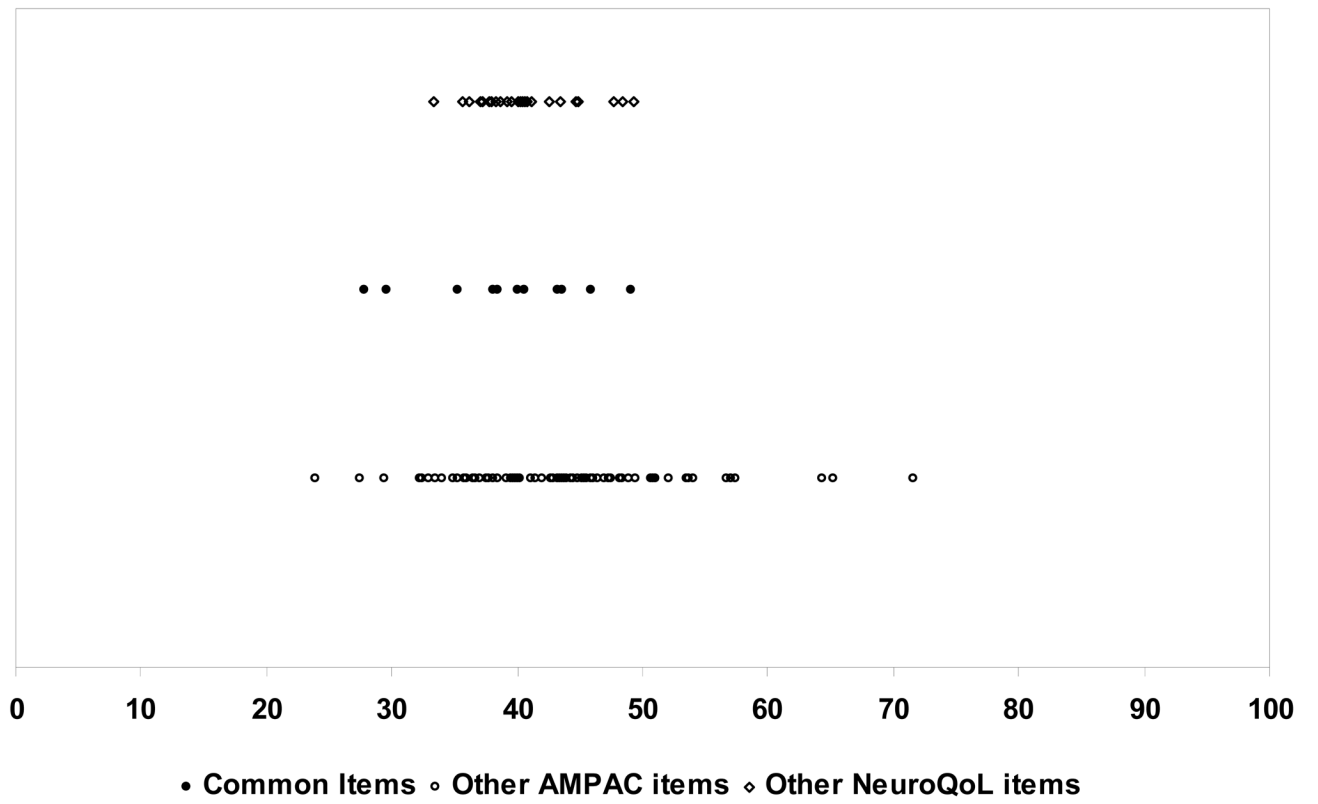
Figure 2.
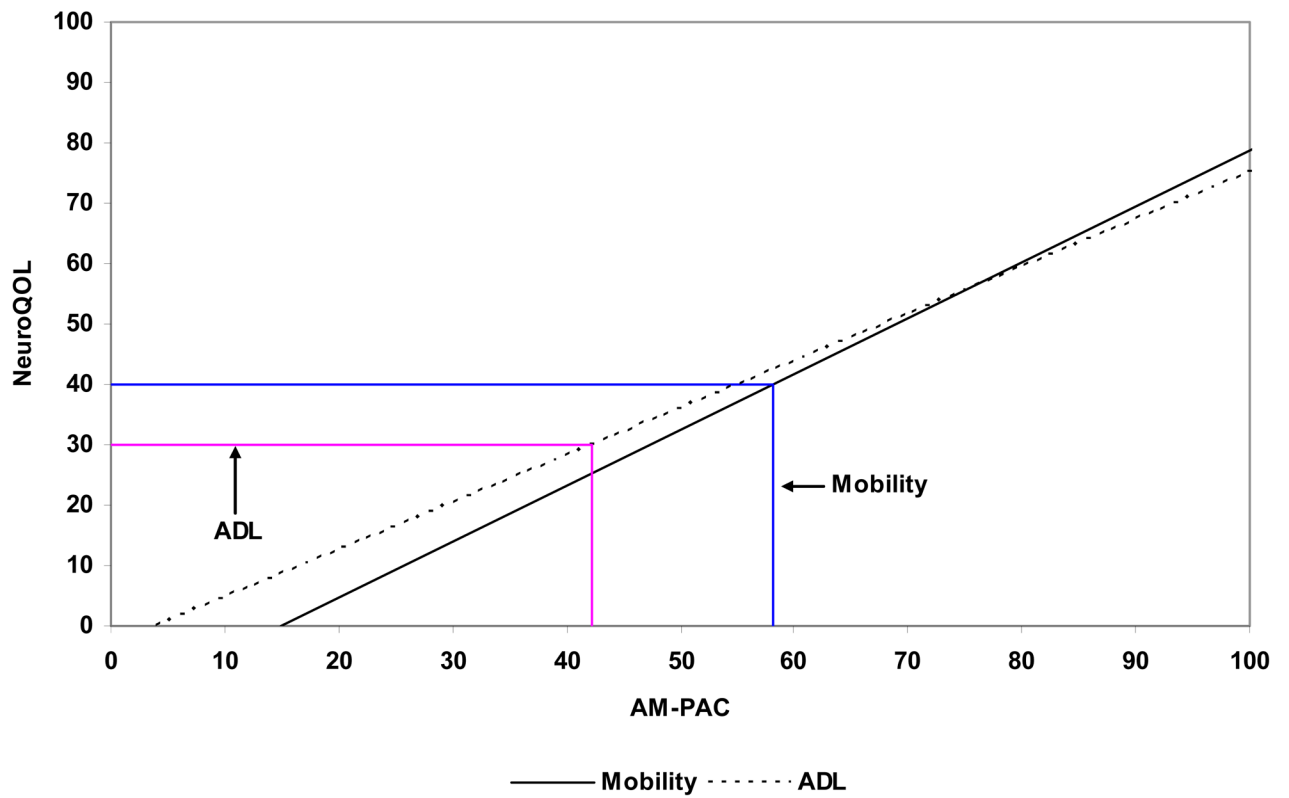Item Map of ADL Items (Common items and remaining AM-PAC and NeuroQOL)

Figure 3.
Scaled score (Mean=50; Sd=10) conversions for AM-PAC and NeuroQOL Mobility and ADL scales

Note: Linear conversions formulae are as follows

Mobility: $\theta_{AM-PAC} = 1.08 \times \theta_{NeuroQOL} + 14.9$

Daily Activity: $\theta_{AM-PAC} = 1.28 \times \theta_{NeuroQOL} + 3.9$
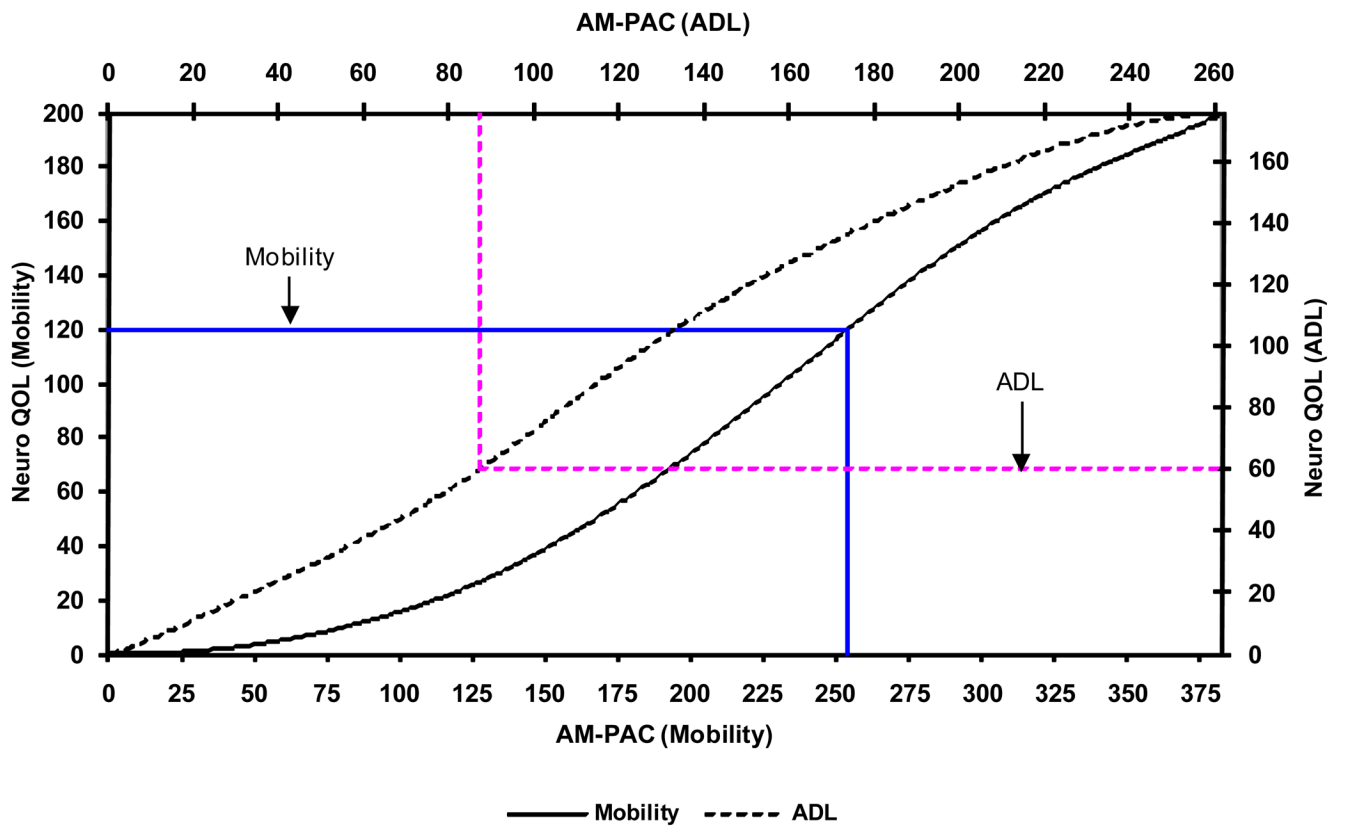
Figure 4.
Raw Score Conversions for AM-PAC and NeuroQOL Mobility and ADL Scales:

Table 1

Summary of Demographics for AM-PAC and NeuroQOL Samples

	AM-PAC Sample	NeuroQOL Sample
N	1041	549
Mean Age (SD)	63.6 (16.6)	52.9 (14.7)
Gender (% female)	58.6 %	51%
Race (Caucasian)	86.5%	91%
Ethnicity (Hispanic)	3%	6%
Married	39.1%	53%

Table 2

Linking Coefficients

Scale	Linking coefficients	Value	Standard Error[*]	Z value(P)
Mobility	A	1.08	0.007	11.43(P<.05)
	B	1.89	0.008	236.25(p<.05)
ADL	A	1.28	0.024	11.67(<.05)
	B	1.79	0.037	48.31(<.05)

[*] Calculated based on the delta method; sample size equals 500