



Published in final edited form as:

Arch Phys Med Rehabil. 2008 June ; 89(6): 1046–1053. doi:10.1016/j.apmr.2007.11.036.

An Exploratory Analysis of Functional Staging Using an Item Response Theory Approach

Wei Tao, PhD Candidate, Stephen M. Haley, PhD, PT, Wendy J. Coster, PhD, OTR, Pengsheng Ni, MD, MPH, and Alan M. Jette, PhD, PT

From the Health and Disability Research Institute, School of Public Health, Boston University Medical Center, Boston, MA (Tao, Haley, Ni, Jette); and Department of Occupational Therapy and Rehabilitation Counseling, Sargent College of Health and Rehabilitation Sciences, Boston University, Boston, MA (Coster).

Abstract

Objectives—To develop and explore the feasibility of a functional staging system (defined as the process of assigning subjects, according to predetermined standards, into a set of hierarchical levels with regard to their functioning performance in mobility, daily activities, and cognitive skills) based on item response theory (IRT) methods using short-forms of the Activity Measure for Post-Acute Care (AM-PAC); and to compare the criterion validity and sensitivity of the IRT-based staging system to a non-IRT-based staging system developed for the FIM instrument.

Design—Prospective, longitudinal cohort study of patients interviewed at hospital discharge and 1, 6, and 12 months after inpatient rehabilitation.

Setting—Follow-up interviews conducted in patients' homes.

Participants—Convenience sample of 516 patients (47% men; sample mean age, 68.3y) at baseline (retention at the final follow-up, 65%) with neurologic, lower-extremity orthopedic, or complex medical conditions.

Interventions—Not applicable

Main Outcome Measures—AM-PAC basic mobility, daily activity, and applied cognitive activity stages; FIM executive control, mobility, activities of daily living, and sphincter stages. Stages refer to the hierarchical levels assigned to patient's functioning performance.

Results—We were able to define IRT-based staging definitions and create meaningful cut scores based on the 3 AM-PAC short-forms. The IRT stages correlated as well or better to the criterion items than the FIM stages. Both the IRT-based stages and the FIM stages were sensitive to changes throughout the 6-month follow-up period. The FIM stages were more sensitive in detecting changes between baseline and 1-month follow-up visit. The AM-PAC stages were more discriminant in the follow-up visits.

Conclusions—An IRT-based staging approach appeared feasible and effective in classifying patients throughout long-term follow-up. Although these stages were developed from short-forms, this staging methodology could also be applied to improve the meaning of scores generated from IRT-based computerized adaptive testing in future work.

Keywords

Outcome assessment (health care); Rehabilitation

Recently, there has been an upsurge of interest in the use of item response theory (IRT) methods to develop and validate health care outcome measures in rehabilitation and related fields.¹⁻⁷ IRT methods consist of a family of scaling models that can be used to calibrate items onto a common metric, thus developing an order of performance (ability levels) of items within a specific outcome area.^{5,8-11} Once the structure and ordering of items is determined, items can be selected for item banks that can be used for a variety of short-form or computer adaptive testing (CAT) applications. A unique feature of this approach is that regardless of which items are selected for a short-form or CAT application, each form is scored on a similar metric, allowing for scoring comparability across short-forms within the entire item bank as well as with CAT versions constructed from the same item bank.

IRT models generate linear person summary scores for each functional domain, the construct to be measured by a scale. The domain scores can then be transformed from logit units to any desired scale, such as a T score with mean of 50 and standard deviation (SD) of 10, or 0 to 100 score. However, 1 challenge with the IRT scaled score is the interpretation and implications of these numeric values. We know that in IRT each domain represents 1 unidimensional ability continuum, ranging from negative infinity to positive infinity. Patients placed at the higher end of the continuous scale have higher functional ability than those who are at the lower end. But the score itself does not exemplify what the patients can or cannot accomplish and thus is less useful in assisting clinicians to make decision about prognosis, placement, or further care. In this regard, classifying patients into different hierarchic stages on each domain may be a more meaningful and interpretable alternative. Stages can be understood as a number of hierarchic levels along the ability continuum that summarize similar physical functioning at the same stage and distinguish different features across stages. Staging refers to the process in which patients are assigned into stages. We have found that using item maps,¹² a display of the pattern of expected functional item difficulties, and developing cut-points to define stages for a pediatric population¹³⁻¹⁶ has been useful at the individual clinical level. Others have also used the concept of an item map to improve the interpretability of their instruments.¹⁷⁻²⁰ This exploratory analysis was undertaken to demonstrate the merits of a staging method to interpret the implications of numeric values derived from IRT scaled scores that are now becoming common in rehabilitation and postacute care clinical instruments.

Functional staging applications in rehabilitation medicine intend to provide a standardized shorthand expression of patient function in various domains of related activities. the Functional Independence Staging (FIS), based on the 18-item FIM defines a stage for each of the 7-performance levels of the FIM (independent to totally dependent) for 4 content domains: (1) activities of daily living (ADLs), (2) sphincter management, (3) mobility, and (4) executive control.²¹ To be classified into a stage in a domain, the patient must function at or above the ratings specified for each activity in that domain. Patient's stages are expected to have prognostic significance and to be useful in the selection of alternative rehabilitation therapies, assistive technologies, or environmental modifications.²² It appears that the FIS has potential to be used in a skilled nursing home setting to both predict outcomes and future patient needs.²³

The approach used in the FIS is not applicable for IRT-based item-bank measures (short-form and CAT applications) such as the Activity Measure for Post-Acute Care (AM-PAC)²⁴ because the FIS is based on individual item raw scores from the FIM, while the AM-PAC employs an IRT model which generates continuous global domain scores. In an IRT model,

individual item raw scores only contribute to the estimation process rather than score reporting. Moreover, it would be impractical to define the AM-PAC staging at each individual item level because the number of items in each AM-PAC domain is much larger than that of the FIS. However, item maps can be used as the raw data to determine meaningful stages and cut-points. The purpose of this study was to develop a prototype staging system for the short-form version of the AM-PAC²⁵ by: (1) developing staging definitions for each of the 3 content domains, the number of stages within a domain, and the expected performance of patients in each stage; (2) deciding cut scores on the continuous item-map scales for each domain to classify patients into stages as defined in the previous step; and (3) evaluating criterion validity and sensitivity to change of the AM-PAC staging system. FIM stages were also assigned to the sample using FIM scores that were part of the dataset.²⁶ Parallel analyses on criterion validity and sensitivity to change were conducted on the FIS and compared with the results of the prototype AM-PAC staging system.

METHODS

Participants

Participants (N=516) were adults aged 18 years and older from the rehabilitation outcomes study who were recruited at discharge from a large inpatient acute care hospital or on admission at one of 2 large inpatient rehabilitation facilities in the greater Boston, MA, region. All subjects had a primary diagnosis of neurologic disorder, orthopedic condition (fractures, joint replacements), or medically complex conditions. Specific inclusion criteria were as follows: patients were currently receiving and/or about to be referred to skilled rehabilitation services; able to speak and understand English; and have a prognosis for survival of 1 year, as determined by the primary physician or a facility recruiter through medical record review. Specifically, the presence of any of the following criteria indicated ineligibility: any orientation deficit, difficulty remembering the day's events, receptive or expressive communication deficits that precluded the patient from communicating responses reliably (verbally or nonverbally). Specific details of the sample are reported in Coster et al.²⁶ Study procedures were approved by the Boston University Institutional Review Board as well as the research review committees of the participating institutions.

Data Collection

The FIM was completed prior to inpatient discharge by clinician report or in some cases by clinician interview if the FIM was not administered routinely at hospital discharge (baseline), and then was collected in the home setting by patient interviews at the 1-, 6-, and 12-month visits. AM-PAC data were collected through clinician interview at baseline and at the same time intervals as the FIM (1, 6, 12mo). Patient interviews were conducted by trained interviewers at the subject's current living location or at a mutually convenient location. A window of 6 weeks from the due date to be interviewed was applied. Subjects not interviewed within this time interval were dropped from that time point. Each interview lasted between 45 and 60 minutes.

The AM-PAC—The initial content domains and item definition for the AM-PAC item pool were guided by the World Health Organization's *International Classification of Functioning, Disability and Health*²⁷ definition and categories of activity. Subsequent factor analyses and Rasch analyses of data from a sample of over 400 persons receiving rehabilitation services led to the definition of 3 separate activity scales.²⁴ The basic mobility domain includes 120 basic physical activities such as bending, walking, carrying, or climbing stairs. The daily activity domain encompasses 65 distinct personal care and instrumental activities; and the applied cognitive domain contains 47 items involving functional application of cognitive skills. Adequate levels of reliability of individual items and validity of the AM-PAC have been

established and are reported previously.^{24,28} Coverage range, unidimensionality, reliability, and validity of the 3 activity scales in AM-PAC were confirmed in subsequent analyses.^{24, 29,30} Each of the 3 AM-PAC short-form scales consists of 10 items that ask about either the difficulty (5-point rating) or use of assistance (6-point rating) to perform specified daily activities. Subjects were given a response card with the relevant response options in large print to use during this part of the interview. In this study, subjects were administered either the hospital version of the AM-PAC short-form if in the inpatient facility, or the community form if interviewed at home.²⁵ The community form of the AM-PAC short form, which includes both basic ADLs (eg, completing grooming activities) and activities more typically performed at home, such as walking several blocks, putting dishes away, or looking up a telephone number was used for home interviews. The community form is linked to an inpatient version of the AM-PAC, which contains only activities likely to be performed in that setting. The linked format supports continuous tracking of a patient across the full spectrum of settings using a single scale. The items for each version were selected from the AM-PAC item pool by scale developers in such a way that item difficulties were appropriate to the population being measured in the facility or community settings. The 3 activity scales were derived from Rasch analyses conducted on the item pool therefore the scores are interval-level data (see details elsewhere²⁵). They range from 0 to 100 with higher scores reflecting greater function (less difficulty, less use of assistance).

FIM instrument—FIM discharge data, when available, was extracted from the medical records or patient charts. In cases in which FIM was not routinely administered data collectors used the patient interview (telephone) version. The patient-interview version of the FIM used for the follow-up interviews has been tested in a similar population using both phone and in-person interview methods with acceptable reliability of the resulting scores.³¹ The FIM contains 18 activity items, each having 7 response options from 1 to 7; the higher the number, the more independent a patient is in performing functional tasks. There are 6 items in the ADLs domain, 2 in the sphincter domain, 5 in the mobility domain, and 5 in the executive control domain. For each domain, 7 stages were defined empirically to mirror the 7 performance levels of the FIM.²¹ To be classified into a FIM stage in a certain domain, the patient must function at or above the predefined ratings of all FIM items in that domain.²¹ The basic mobility, daily activity, and applied cognitive domains in the AM-PAC are analogous to the mobility, ADLs, and executive control domains in the FIM, respectively. The sphincter management domain in the FIM has no AM-PAC equivalent.

Analytic Strategy

We applied techniques used successfully in educational testing to develop a staging plan for the AM-PAC. Staging in rehabilitation is similar to setting performance standards in education, which is defined as the process by which a standard or cut score is established.³²

Standards—We developed an initial set of standards that specified the anticipated number of stages a domain and the expected performance of patients in each stage. We decided on no more than 5 discrete stages for each domain, and discussed some general stage content expectations for each stage within a domain. The decision was based on an empirical examination on the content and difficulty levels of each individual item, and to create a staging system that was both succinct enough to summarize similar performances in the same stage and distinctive enough to discriminate differences across stages.

Cut-scores—We adopted a technique called the Bookmark³² method to decide on the cut-scores for stages in each AM-PAC domain, and modified it for this exploratory analysis. The Bookmark method is 1 of several item-mapping procedures developed in an attempt to simplify the cognitive task of standard setters. This method has been widely used in K through 12

educational settings since it was first introduced in 1996.³² Advantages of the Bookmark method include the capability of setting up multiple cut scores, analyzing polytomously scored items and using item calibrations estimated from the IRT analyses to facilitate the judgment.

To implement the Bookmark method, a panel of judges from the measurement field and content field are formed. Each judge is provided with 2 crucial documents—the preset 5-stage standards and an ordered item-category list. The 5-stage standards lay out the expected performance of patients in each stage. The ordered item-category list contains item category estimates sorted from the easiest to the hardest across all items. Each item appears exactly $m - 1$ times, once for every item-category-threshold (m is the number of response categories in each item). For each item category, the scale score corresponding to a .50 probability of endorsing the current and lower item categories is computed.

Based on the standards and using the ordered item-category list, judges put a bookmark on the item-category-threshold if they believe that patients at the current stage are very unlikely (have a the probability of <50%) to achieve the performance described by the higher categories above the threshold.³³ The corresponding scale score is the cut point for 2 adjacent stages. Because this is an exploratory analysis, rather than conduct a formal bookmark procedure using multiple judges examining item threshold values for each item, the first author examined each item difficulty and threshold and with feedback from other authors, provided suggested cutoff scores for each adjacent stage.

We used person scores from the longitudinal data to estimate subject locations on the AM-PAC item map. Item parameter estimates in the item map came from a previous calibration study.²⁵ Because this was an exploratory analysis and the Bookmark method does not need to consider every item in the bank, we employed a selection of AM-PAC items that have been used as a shorter substitute for the full item bank³⁴ to develop the staging plan. Twenty items from the basic mobility domain, 16 items from the daily activity domain, and 18 items from the applied cognitive were used as item content. These items are representative of the AM-PAC item pool in terms of the content coverage and difficulty spectrum.

Criterion validity—For each AM-PAC domain and its equivalent FIM domain, we selected 2 criterion variables assumed to be highly related to the construct of what each domain is measuring. These criterion variables came from the Participation Measure of Post-Acute Care³⁵ scale which was collected together with the AM-PAC scale during the 3 home follow-up visits.²⁶

For the FIS executive control and AM-PAC applied cognitive, the 2 criterion variables were: “How much are you currently limited in communicating with others in general?” (communication with others) and “How much are you currently limited in managing your money, such as keeping track of expenses and paying bills?” (managing money). The 2 comparative items for the FIS mobility and AM-PAC basic mobility were: “Thinking about how you go places using any help or means of transportation available, how much are you currently limited in getting around?” (limited in going places) and “How much are you currently limited in getting around at your home?” (getting around home). Two items were used to examine the validity of the AM-PAC daily activity stages and the FIS ADL stages: (1) “Taking into account any help or services that are available to you, how much are you currently limited in providing personal care to yourself and others?” (limited in personal care) and (2) “Taking into account any help or services that are available to you, how much are you currently limited in keeping your home clean and fixed up?” (limited in home clean and fix up). Spearman correlation was calculated between each criterion variable and the AM-PAC domain stages, as well as between the same criterion variable and the FIS.

Sensitivity to change—We created complementary cumulative distribution functions (CDF) plots²² showing the proportion of patients above each stage by interview period. Then, we conducted a global test on the significance of cross-visit score difference using the Friedman test, a nonparametric version of repeated-analysis of variance test (ANOVA) aiming at detecting differences across repeated measures or time points. The Friedman test was chosen over the repeated ANOVA procedure because stages are at an ordinal level rather than an interval level. Finally, any significant Friedman test was followed by a post hoc pairwise comparison between adjacent visits using the Wilcoxon rank-sum test, a nonparametric version of the paired *t* test. The same set of analyses in examining sensitivity to change was repeated on the FIS.

RESULTS

The study sample included slightly more women than men (53% vs 47%) and a greater percentage classified in the complex medical category (44%) compared with the lower-extremity orthopedic (32%) and neurologic (24%) categories. The mean age of participants was 68.3 years, however, the range extended from 19 to 100 years with about 20% of subjects younger than age 50 years. At follow-up, 417 (81%) of participants were seen at 1 month, 370 (72%) were interviewed at 6 months, and 336 (65%) were seen again at 12 months. For analyses involving change across time periods, the number of subjects varied from the total given above because some subjects were missing data for 1 of the follow-up assessments. See Coster et al²⁶ for more details on the sample characteristics.

To detect possible differences in demographic characteristics and disease severity between those who completed the study (completers) and those who missed at least 1 visit (dropouts), we examined 4 background variables of age, sex, race (white vs nonwhite) and education (high school or below, some college, and college degree or higher), as well as their baseline scores for each AM-PAC scales. Results have shown that completers and dropouts do not differ in age, sex composition, race, and all 3 baseline scale scores, but do differ in their educational levels. Participants with higher educational attainment tended to stay in the study. Specifically, 39% of completers held college degree or higher, while only 20% of dropouts had the same educational level. For the criterion-validity analysis, pairwise deletion was employed to calculate the Spearman correlation while for the sensitivity analysis, list-wise deletion was employed to include those who completed the study.

AM-PAC Stage Definitions

A performance description and final score range for each of the 5 stages are presented in table 1. For domains of daily activity and applied cognitive, the expected performances of stage-specific patients are differentiated across content coverage. For example, in the applied cognitive domain, stages are differentiated by the fact that oral communication is relatively easier than written communication and other complicated tasks involving cognitive skills. For the basic mobility domain, the stages are defined by the patient's radius of movement, such as whether patients can move freely within 1 room, within a building, outside of the building, or outside of a building doing active sport exercises.

Counts of patients classified in each stage, shown in table 2, reveal a general pattern that the number of patients defined at lower stages (stage 1, 2, and 3) decreased from the baseline visit to the 12-month visit, while this number increased for patients at higher stages (stages 4 and 5), which implies that in general more patients showed improved functional performance in all three domains during the follow-up period.

Table 3 lists the percentage of patients at or below stage 4 who have improved at least 1 stage between adjacent visits. The largest improvement occurred between the baseline visit and the

1-month follow-up, with more than 50% of patients showing improvement. The smallest change occurred between the 6- and 12-month follow-up visits, with about 20% to 30% patients showing improvement.

Criterion Validity

Spearman correlation between each of the criterion variables and AM-PAC domain stages as well as with the FIS is shown in table 4. Almost all AM-PAC domains significantly correlated with corresponding criterion variables with the basic mobility domain having the highest correlation coefficients ($>.40$). The FIS showed a similar pattern, but the magnitude of the correlation coefficients were smaller than that of the AM-PAC stages.

Sensitivity to Change

The complementary CDF plot on the AM-PAC staging and the FIS are shown in figures 1A and 1B. In the chart, the x axis represents stage value, the y axis represents the percentage of patients who score at or above each stage, and each line represents a follow-up visit. If patients functioned better during a later visit, more patients would be defined at a higher stage and thus the percentage of patients above each stage would be higher. In this case, we would expect to see the later-visit line lie above the former-visit line. The larger the gap between 2 lines, the greater the stage change between the 2 adjacent visits.

Both staging systems are more sensitive to changes from baseline to the 1-month follow-up visit than all other follow-up visits. The change pattern of the applied cognitive domain in AM-PAC is very similar to the pattern of the executive control domain in the FIS. When comparing the basic mobility and daily activity domains in the AM-PAC to corresponding domains in FIS, we see gaps between the baseline and 1-month visit are larger in FIS than in the AM-PAC staging system suggesting that FIS may be more sensitive to changes during these 2 time points. This is probably due to the fact that fewer stages were defined in the AM-PAC than in the FIS. However, the slopes of time lines in the FIS is flatter than those in the AM-PAC for the 3 follow-up visits between stages 1 and 5, which indicates these stages in the FIS could not differentiate patients very well at the later follow-up periods.

Table 5 presents the sensitivity analysis results. The global Friedman test revealed that both staging systems were able to detect overall changes across the 4 test intervals. Further post hoc pairwise comparisons between adjacent time points suggested that significant changes occurred between the baseline to the 1-month visit, as well as from 1- to 6-month visits. However, it should be noted that for the 1- to 6-month comparison, AM-PAC detected a positive change in the applied cognitive domain, while FIS detected a negative change in the executive control domain. AM-PAC detected no significant change between the last 2 visits (6–12mo), while FIS detected a significant negative change in the executive control domain, which is due to the fact that a large percentage of people were at ceiling on the FIM executive control domain at baseline and this proportion declined over time.

DISCUSSION

To date, the rehabilitation and postacute care industries have limited experience and success achieving a standardized, patient-centered outcome assessment approach that can provide interested stakeholders with appropriate information on outcomes, quality of care, and a profile that highlights specific patterns of disability. Outcome systems will need to provide information that is interpretable to both clinicians and patients, that can be applied over time and across different settings, and that can predict important health outcomes. The latter is increasingly important to help patients and clinicians in their decision-making, care planning, and in

improving health care services.³⁶ To date, only the FIS²² achieves many of these goals, however, its appropriateness across a wide variety of postacute care settings may be limited.

We have demonstrated that an IRT-based approach toward developing a staging system with the AM-PAC is both feasible and psychometrically sound, and compares favorably to the existing FIS. A major advantage of the AM-PAC staging system at first glance is that the content provides a broader perspective of patient abilities than the FIM. Another advantage of the IRT-based approach may be that it can be used for the many new instruments that are being developed with an underlying assumption of item order and fit.

There are both similarities and differences between the FIS and the proposed IRT-based staging systems. In the case of FIS, the stages are defined by the most likely set of items within a domain that are seen in the extensive FIM database. For example, stage 6 (modified independence) is chosen if all the items in the mobility domain except stair climbing are at a modified independent level. A relative advantage of the FIM staging system is that once a stage is passed the clinician is certain that the patient is able to function at or above the specified level for each component activity. Consequently a minimum level of ability or function is a guarantee for each item making clinical decision-making relatively certain.

In the case of the AM-PAC staging, an overall hierarchic model of items is developed from the data and is fit to express the most common sequence and location of items along a continuum of difficulty. Stage boundaries are then defined based on the pattern of the data and on input from clinical experts. Stages are defined according to the most likely levels of item functioning within the specified score range. The patient's actual status on the activities may not be known, but rather is inferred through an empirically derived probability. These inferences, however, can be checked by examining actual versus expected performance. Both the FIS and AM-PAC staging systems cover similar mobility, ADLs, and cognitive-based items, while the FIS also includes sphincter control.

The prototype AM-PAC stages appeared to describe the recovery patterns of the postacute care sample, had good correspondence with key mobility, ADLs, and cognitive criterion items, and showed sensitivity to changes throughout the entire 12-month follow-up period. Because the AM-PAC has broader and more comprehensive functional content than the FIM, the data suggest it may be more useful as a long-term monitoring system at 6- and 12-months posthospital discharge. In future studies, however, we will need to examine more closely how patients shift stages both in the positive and negative directions at later follow-up periods, and whether these variations can be explained by the instrument or staging methods used, potential ceiling effects, or other artifacts of collecting repeated data from individuals.

Study Limitations

This exploratory study has several limitations. We did not fully implement the Bookmark procedure using many judges, and we recognize that our stage definitions and cut-points might have been somewhat different had the full procedure been implemented. We also limited the AM-PAC item content examined for the staging cut-point decisions to about one third of the full AM-PAC item bank. In a more rigorous procedure, we would examine much more item content when making cut-point decisions. Finally, problems of missing data, as any longitudinal study is likely to encounter, may limit generalization of the study results. However, in this study, we found no relationship between the status of dropouts and patient's baseline severity—the construct being measured—which, to some degree, indicates that the difference between the dropouts and the completers may have less impact on the findings.

CONCLUSIONS

This study developed a staging plan for the IRT-based AM-PAC domains. Five stages were defined for each AM-PAC domain. The criterion validity of the AM-PAC stages was supported by results indicating that all AM-PAC domain stages significantly correlated with external variables assumed to be highly related to the construct of each AM-PAC domain. AMPAC stages had a higher correlation with these criterion variables than the corresponding FIS domains. The AM-PAC staging system was as sensitive as the FIS in detecting changes between adjacent visit points throughout the follow-up visits, but was less sensitive than FIS in detecting the baseline to 1-month change. The results of this exploratory study suggest that IRT-based methods may be appropriate for creating staging systems for the new generation of short-form and CAT measures with well-defined item banks that are currently being developed or are already in use.

Acknowledgements

Supported by the National Institute of Disability and Rehabilitation Research, U.S. Department of Education (grant no. H133B990005); National Institute of Child Health and Human Development and the Agency for Healthcare Research and Quality (grant no. R01 HD043568); and an Independent Scientist Award (grant no. K02 HD45354-01).

A commercial party having a direct financial interest in the results of the research supporting this article has conferred or will confer a financial benefit upon the author or one or more of the authors. Haley and Jette have a stock interest in CRE Care, which distributes the Activity Measure for Post-Acute Care products.

References

1. Jette AM, Haley SM. Contemporary measurement techniques for rehabilitation outcomes assessment. *J Rehabil Med* 2005;37:339–345. [PubMed: 16287664]
2. Cella D, Gershon R, Lai JS, Choi S. The future of outcomes measurement: item banking, tailored short forms, and computerized adaptive assessment. *Qual Life Res* 2007;16:133–141. [PubMed: 17401637]
3. Fries J, Bruce B, Cella D. The promise of PROMIS: using item response theory to improve assessment of patient-reported outcomes. *Clin Exp Rheumatol* 2005;23:S53–S57. [PubMed: 16273785]
4. Cella D, Yount S, Rothrock N, et al. The Patient-Reported Outcomes Measurement Information System (PROMIS): progress of an NIH Roadmap Cooperative Group during its first two years. *Med Care* 2007;45(5 Suppl):S3–S11. [PubMed: 17443116]
5. 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: Cambridge Univ Pr; 2005. p. 445-464.
6. Fayers P. Applying item response theory and computer adaptive testing: the challenges for health outcomes assessment. *Qual Life Res* 2007;16:187–194. [PubMed: 17417722]
7. Ware JE Jr, Gandek B, Sinclair SJ, Bjorner B. Item response theory in computer adaptive testing: implications for outcomes measurement in rehabilitation. *Rehabil Psychol* 2005;50:71–78.
8. Hambleton, R.; Swaminathan, H. Item banking. In: Hambleton, R.; Swaminathan, H., editors. *Item response theory: principles and applications*. Boston: Kluwer Nijhoff; 1985. p. 255-279.
9. Revicki DA, Cella DF. Health status assessment for the twenty-first century: item response theory, item banking and computer adaptive testing. *Qual Life Res* 1997;6:595–600. [PubMed: 9330558]
10. Bode RK, Lai JS, Cella D, Heinemann AW. Issues in the development of an item bank. *Arch Phys Med Rehabil* 2003;84:S52–S60. [PubMed: 12692772]
11. Hays RD, Morales LS, Reise SP. Item response theory and health outcomes measurement in the 21st century. *Med Care* 2000;38(9 Suppl):II28–II42. [PubMed: 10982088]
12. Coster W, Ludlow L, Mancini M. Using IRT variable maps to enrich understanding of rehabilitation data. *J Outcome Meas* 1999;3:123–133. [PubMed: 10204323]

13. Haley, SM.; Coster, WJ.; Ludlow, LH.; Haltiwanger, JT.; Andrellos, PA. Pediatric evaluation of disability inventory: development, standardization and administration manual. Boston: Trustees of Boston Univ; 1992.
14. Fragala MA, Haley SM, Dumas HM, Rabin JP. Classifying mobility recovery in children and youth with brain injury during hospital-based rehabilitation. *Brain Inj* 2002;16:149–160. [PubMed: 11839109]
15. Dumas H, Haley S, Bedell G, Hull EM. Social function changes in children and adolescents with acquired brain injury during inpatient rehabilitation. *Pediatr Rehabil* 2001;4:177–185. [PubMed: 12160358]
16. Dumas H, Haley S, Fragala MA, Steva BJ. Self-care recovery of children with brain injury: descriptive analysis using the Pediatric Evaluation of Disability Inventory (PEDI) functional classification levels. *Phys Occup Ther Pediatr* 2001;21:7–27. [PubMed: 12029856]
17. Malec JF, Moessner AM, Kragness M, Lezak MD. Refining a measure of brain injury sequelae to predict postacute rehabilitation outcome: rating scale analysis of the Mayo-Portland Adaptability Inventory. *J Head Trauma Rehabil* 2000;15:670–682. [PubMed: 10745183]
18. Bode RK, Heinemann AW, Semik P. Measurement properties of the Galveston Orientation and Amnesia Test (GOAT) and improvement patterns during inpatient rehabilitation. *J Head Trauma Rehabil* 2000;15:637–655. [PubMed: 10745181]
19. Ryser L, Wright B, Aeschlimann A, Mariacher-Gehler S, Stucki G. A new look at the Western Ontario and McMaster Universities Osteoarthritis Index using Rasch analysis. *Arthritis Care Res* 1999;12:331–335. [PubMed: 11081002]
20. Stelmack J, Szlyk JP, Stelmack T, et al. Use of Rasch person-item map in exploratory data analysis: a clinical perspective. *J Rehabil Res Dev* 2004;41:233–242. [PubMed: 15558377]
21. Stineman MG, Ross RN, Fiedler R, Granger CV, Maislin G. Functional independence staging: conceptual foundation, face validity, and empirical derivation. *Arch Phys Med Rehabil* 2003;84:29–37. [PubMed: 12589617]
22. Stineman MG, Ross RN, Fiedler R, Granger CV, Maislin G. Staging functional independence validity and applications. *Arch Phys Med Rehabil* 2003;84:38–45. [PubMed: 12589618]
23. Jette DU, Warren RL, Wirtalla C. Validity of functional independence staging in patients receiving rehabilitation in skilled nursing facilities. *Arch Phys Med Rehabil* 2005;86:1095–1101. [PubMed: 15954046]
24. Haley SM, Coster WJ, Andres PL, et al. Activity outcome measurement for post-acute care. *Med Care* 2004;42:I49–I61. [PubMed: 14707755]
25. Haley SM, Andres PL, Coster WJ, Kosinski M, Ni PS, Jette A. Short-Form Activity Measure for Post-Acute Care. *Arch Phys Med Rehabil* 2004;85:649–660. [PubMed: 15083443]
26. Coster WJ, Haley SM, Jette AM. Measuring patient-reported outcomes after discharge from inpatient rehabilitation settings. *J Rehabil Med* 2006;38:237–242. [PubMed: 16801206]
27. World Health Organization. International classification of functioning, disability and handicap: ICF. Geneva: WHO; 2001.
28. Andres PL, Haley SM, Ni PS. Is patient-reported function reliable for monitoring post-acute outcomes? *Am J Phys Med Rehabil* 2003;82:614–621. [PubMed: 12872018]
29. Coster WJ, Haley SM, Andres PL, Ludlow LH, Bond TL, Ni PS. Refining the conceptual basis for rehabilitation outcome measurement: personal care and instrumental activities domain. *Med Care* 2004;42:I62–I72. [PubMed: 14707756]
30. Coster WJ, Haley SM, Ludlow LH, Andres PL, Ni PS. Development of an applied cognition scale to measure rehabilitation outcomes. *Arch Phys Med Rehabil* 2004;85:2030–2035. [PubMed: 15605343]
31. Petrella RJ, Overend T, Chesworth B. FIM after hip fracture: is telephone administration valid and sensitive to change? *Am J Phys Med Rehabil* 2002;81:639–644. [PubMed: 12172514]
32. Cizek G, Bunch M, Koons H. Setting performance standards: contemporary methods. *Educ Meas Iss Pract* 2004;23:31–50.
33. Huynh H, Meyer J. Maximum information approach to scale description for affective measures based on the Rasch model. *J Appl Meas* 2003;4:101–110. [PubMed: 12748403]

34. Haley S, Siebens H, Coster W, et al. Computerized adaptive testing for follow-up after discharge from inpatient rehabilitation: I. Activity outcomes. *Arch Phys Med Rehabil* 2006;87:1033–1042. [PubMed: 16876547]
35. Gandek B, Sinclair J, Jette A, Ware J. Development and initial testing of the Participation Measure for Post-Acute Care (PM-PAC). *Am J Phys Med Rehabil* 2007;86:57–71. [PubMed: 17033591]
36. Jette AM, Haley SM, Ni PS. Comparison of functional status tools used in post-acute care. *Health Care Financ Rev* 2003;24:13–24. [PubMed: 12894632]

Figure 1a: AM-PAC Staging

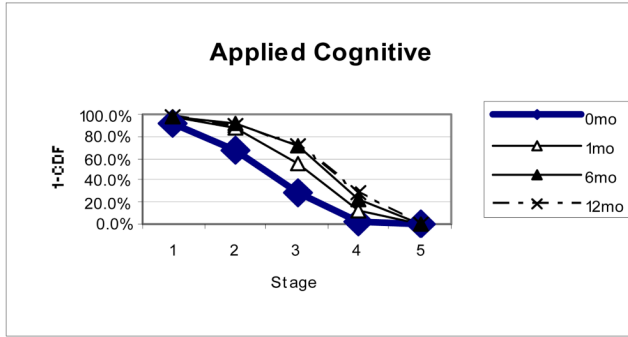


Figure 1b: FIM Staging

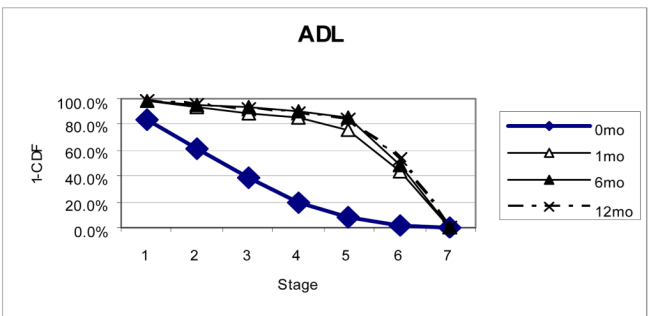
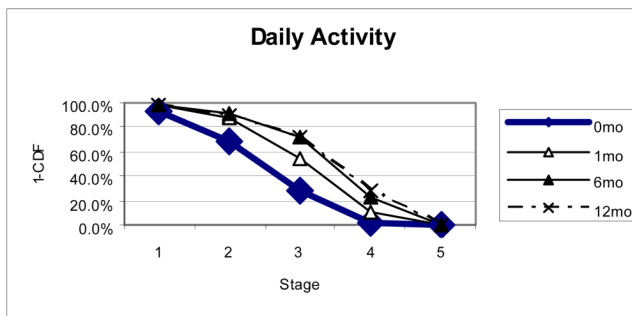
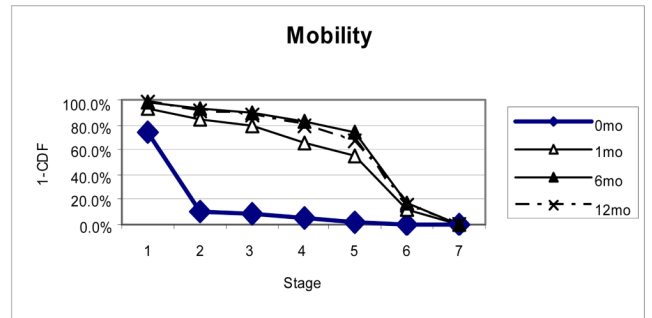
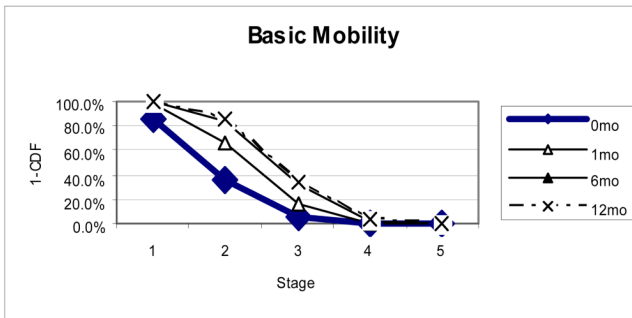
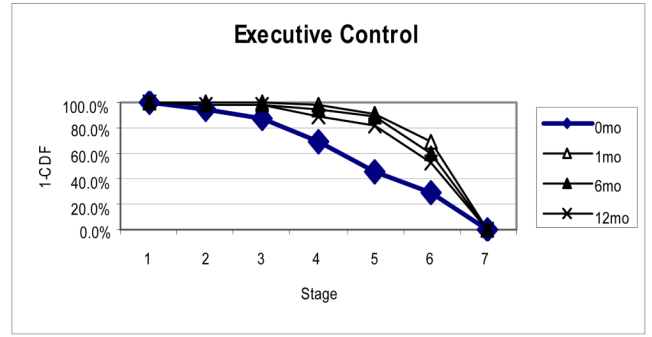


Figure 1.
Complementary CDF Plot for the AM-PAC and FIM staging systems

Table 1
Stage Descriptions of Activity Measure for Post-Acute Care (AM-PAC)

AM-PAC Applied Cognitive Stages		
AM-PAC Stage	Score Interval	Stage Description
1	0–44	Unable to conduct written communications and any task involving cognitive skills; unable or has difficulty with oral communication.
2	45–52	Some difficulty in oral communication with others; unable or has difficulty in reading or conducting any complicated tasks.
3	53–64	Oral communication with a little difficulty; some difficulty in reading and conducting complicated tasks
4	65–88	No difficulty in oral communication or in reading; a little difficulty in managing complicated tasks.
5	89–100	No difficulty in oral or written communication or in conducting complicated tasks involving cognitive skills.
AM-PAC Daily Activity Stages		
1	0–41	Unable to dress, eat and take care of personal grooming activities
2	42–53	Some difficulty in eating, dressing and grooming; unable or lot of difficulty in lower body dressing, bathing and instrumental activities.
3	54–62	Little difficulty in eating, dressing and grooming; some difficulty in lower body dressing, bathing and instrumental activities.
4	63–84	Able to eat, dress and groom his/her self, but a little difficulty in lower body dressing, bathing and instrumental activities.
5	85–100	Independent in their daily activities.
AM-PAC Basic Mobility Stages		
1	0–34	Independent in their daily activities.
2	35–52	Limited mobility inside of a building; unable to do bending/reaching activities.
3	53–66	Little difficulty in moving inside a building, but limited in going outdoors.
4	67–84	Walks independently inside and outside, some difficulty in doing moderate or strenuous activities.
5	85–100	Moves inside or outside independently and participates in strenuous sports.

Table 2

Counts of patients in each AM-PAC stage, by visit

stage	Applied Cognitive visits			Basic Mobility visits			Daily Activity visits							
	0mo	1mo	6mo	0mo	1mo	6mo	0mo	1mo	6mo	12mo				
1	3	2	1	2	2	69	9	1	1	2	39	6	6	5
2	48	13	6	14	14	247	125	55	44	44	124	48	25	28
3	237	73	56	36	36	146	197	179	159	159	199	132	75	62
4	167	202	134	133	133	28	62	101	98	98	128	180	182	146
5	48	124	172	149	149	2	1	9	11	11	13	47	82	94
Total	503	414	369	334	334	492	394	345	314	314	503	413	370	335

Table 3

For patients at or below Stage 4, the percentage of those who improved at least one AM-PAC stage between adjacent visits

Domain	Adjacent Visits		
	0 – 1 month	1 – 6 month	6 – 12 month
Applied Cognitive	59%	50%	32%
Basic Mobility	50%	42%	22%
Daily Activities	54%	40%	25%

Table 4
Spearman correlations between AM-PAC staging, FIS and criterion variables:

	1-mo		6-mo		12-mo	
	Criterion Variable	Criterion Variable	Criterion Variable	Criterion Variable	Criterion Variable	Criterion Variable
AM-PAC –						
Applied						
Cognitive	0.20 [†]	0.17 [†]	0.31 [†]	0.14 [†]	0.42 [†]	0.21 [*]
FIS –						
Executive						
Control	0.16 [†]	–0.03	0.27 [†]	0.03	0.33 [†]	0.08
	Limited in Going Places	Getting Around Home	Limited in Going Places	Getting Around Home	Limited in Going Places	Getting Around Home
AM-PAC –						
Basic Mobility	0.46 [†]	0.55 [†]	0.43 [†]	0.48 [†]	0.49 [†]	0.54 [†]
FIS - Mobility	0.39 [†]	0.39 [†]	0.48 [†]	0.41 [†]	0.43 [†]	0.44 [†]
	Limited in Personal Care	Limited in Home Clean and Fix Up	Limited in Personal Care	Limited in Home Clean and Fix Up	Limited in Personal Care	Limited in Home Clean and Fix Up
AM-PAC –						
Daily						
Activities	0.19 [†]	0.10	0.21 [†]	0.22 [†]	0.27 [†]	0.27 [†]
FIS - ADL	0.13 [*]	0.12 [*]	0.15 [†]	0.14 [†]	0.22 [†]	0.07

* $p < 0.05$

[†] $p < 0.01$

Table 5

Sensitivity to change, AM-PAC and FIS

	AM-PAC Domains			FIS Domains			
	Basic Mobility	Daily Activity	Applied Cognitive	Mobility	ADL	Executive Control	
Friedman test:	215.3 [†]	222.9 [†]	193.1 [†]	271.8 [†]	298.0 [†]	106.7 [†]	
Wilcoxon signed rank sum test on adjacent visits:							
0 – 1 month	4933.0 [†]	7358.0 [†]	6666.0 [†]	7378.5 [†]	9141.0 [†]	4356.5 [†]	
1 – 6 month	2951.0 [†]	2831.0 [†]	2519.0 [†]	3475 [†]	1415.0 [†]	-1129.5 [†]	
6 – 12 month	358.0	190.5	-384.0	-871.0	-377.5	-950.0 [†]	

[†] p<0.01