

Validation of an Item Bank in a Sample of Community-Dwelling Survivors of a Stroke

Chad E. Cook, PT, PhD, MBA, OCS;¹ Adam Goode, PT, DPT;¹; Daniel Erb, PT, PhD;¹
Jan K. Richardson, PT, PhD, OCS;¹ Ricardo Pietrobon, MD, PhD²

¹ Duke University Medical Center, Division of Physical Therapy
² Division of Orthopaedic Surgery, Center for Excellence in Surgical Outcomes, Duke University Health System, Durham, NC

ABSTRACT

Purpose: The purpose of this study was to validate an item bank of questions associated with activities of daily living (ADL), using graded item response theory (G-IRT), in a sample of community-dwelling survivors of a stroke. **Methods:** The study sample consisted of 364 community-dwelling individuals who reported a recent history of stroke. Sixteen line items from the 1999-2000 *National Health and Nutrition Examination Surveys* (NHANES) survey were analyzed using factor analysis, internal analyses of consistency, and G-IRT. **Results:** The 16 line items demonstrated unidimensionality and were internally reliable. Thirteen line items demonstrated good discrimination and suitable thresholds. The majority of items exhibited appropriate sensitivity across the entire spectrum of functional severity indicating that these items closely reflected the relationship of decreased function with increased severity of illness. **Conclusion:** Each of the 13 items is scale independent, is valid for measurement of functional impairment, and may be applicable for use in a scale for assessment of functional change in community-dwelling stroke survivors.

Key Words: stroke, item response theory, activities of daily living, NHANES

INTRODUCTION

Annually, there are an estimated 700,000 recent or recurrent reported strokes that contribute or are an underlying cause of death in an estimated 273,000 individuals.¹ Stroke is a pathology that primarily affects the elderly with risk of occurrence nearly doubling after the age of 55.² The condition is a leading cause of serious long-term disability in the United States,³ which results in residual physical impairments and related dysfunction.⁴

As stroke survival improves, appropriate tests and measures of treatment for stroke-related disabilities are paramount.⁵ Recent advances in the care of stroke have allowed

a greater rate of survivors of stroke who return to the community; although functional outcomes in these individuals often remain limited.⁵ Up to 3 months following the onset of stroke, 17% of survivors remain dependent in personal care.⁵ Up to 6 months following a stroke 50% of survivors are dependent in instrumental activities of daily living (eg, self-administration of medication, meal preparation) and only 33% are independent in community mobility.⁵ Individuals with high levels of recovery continue to have disability related to hand function, performance of basic and instrumental activities of daily living, and physical functioning.⁶⁻⁸

Individuals who survive stroke and receive rehabilitation services generally experience improved short- and long-term survival, higher function, return home more quickly, and may experience an improved quality of life.⁹⁻¹² The use of outpatient rehabilitation, including physical and occupational therapy has been shown to increase the likelihood of further functional improvement.^{5,12} Because survivors of a stroke are living longer and are more commonly integrated into a community-dwelling there has been a trend by rehabilitation specialists, toward the use of patient-report, quantifiable scales that approximate general health and quality of life.¹³

Health related quality of life scales have often explored the relative importance of a range of items on a patient population. Incorporated within these items is physical functioning and/or role activities which are items associated with the ability to perform activities of daily living (ADL).^{14,15} Principal to each patient response scale are validated line items or 'item banks' which represent quantitative changes in ADLs.¹⁶ An item bank is a collection of items for which the measurement properties of each item are known.^{17,8} Therefore, the use of adaptive testing algorithms selects the 'best' items for individual patients for a particular pathology.¹⁶ Item banks have been designed to measure ADL function in mixed neurological patient populations,¹⁶ while indirectly, several ADL function scales have been validated using item response theory, specifically, Rasch analysis.¹⁹⁻²³

Item response theory (IRT) methods are particularly powerful for reducing the problems associated with range and detail, which occur in fixed length scales.²⁴ Current methods using IRT link a person's ability to item difficulty and have solved the problems of 'ceiling' and 'floor' effects of traditional scales.²⁵ Item response theory differs from classical response theory in that it focuses on the validation of individual items, rather than validation of the instrument as a whole.²⁶ Although several forms of IRT exist, graded response IRT (G-IRT) is most appropriate for questions with a Likert type response versus Rasch IRT,

Address correspondence to: Chad Cook PT, PhD, MBA, OCS, Division of Physical Therapy, Duke University, DUMC 3907, Durham, NC 27710, Ph: 919-684-9575 (chad.cook@duke.edu).

which dichotomizes responses and is most appropriate for use in nominal responses.²⁶ Additionally, the line item properties from G-IRT are population independent, thus are less affected by the severity of conditions among different population characteristics.²⁷ Researchers are able to calculate the measurement error of each patient, a method not possible using classic validity methods, which assumes all scores are error free.²⁷ The G-IRT assumes the probability that a respondent will choose an option for the item that is a function of the respondent's underlying trait, illness, or disability,²⁶ and can transform measures of a single latent trait in an associative metric, that is usable in a number of questionnaires.²⁷

The purpose of this study was to create and validate an item bank of questions associated with ADLs, for community-dwelling, survivors of a stroke, using data obtained from the *National Health and Nutrition Examination Surveys* (NHANES).²⁸ The NHANES are an ongoing series of cross-sectional surveys on health and nutrition designed to be nationally representative of the non-institutionalized, US civilian population.²⁸ Each year, nearly 5,000 participants complete the health examination component of NHANES, by answering questions specifically designed to (1) estimate the number and percent of persons in the U.S. with selected diseases and risk factors; (2) monitor trends in the prevalence, awareness, treatment, and control of selected diseases; (3) monitor trends in risk behaviors and environment exposures; (4) analyze risk factors for selected diseases; (5) study the relationship between diet, nutrition, and health; and (6) explore the emerging public health issues and new technologies.²⁸ Using the data within the database, the study endeavors to evaluate the internal reliability, line item validity, and dimensionality of the selected items within the item bank. Because most clinicians do not use ADL scales when treating patients,²⁹ the development of an item bank of questions using IRT allows clinicians to use specific items versus questionnaires during assessment of ADL dysfunction.

METHODS

Subjects

This study used data from the 1999-2000 NHANES.²⁸ The data within NHANES include a household interview and a health examination component.²⁸ The health examination component was performed in a mobile examination center and included an interview, physical examination, and laboratory examination.²⁸ It was during the health examination interview that patients were queried about their history of stroke. Additional information about the weighting and validity of the NHANES survey is available free to the public at the NHANES website.²⁸

Design

Sixteen items (Appendix A) within the NHANES survey that measure activity constructs associated with locomotion and transfers, household productivity, social integration, and manipulation of surroundings, were used to develop an item bank. The items were originally developed by the NHANES

administrators to measure ADLs and were administered to all respondents, regardless of pathology. Because many of participants in the NHANES survey have less than a high school degree, the questions were worded carefully and clearly and were easy to understand.²⁸ Item scores ranged from '1' to '4' and the total score is the summation of all item scores. A Likert-type selection of 1 = no difficulty, 2 = some difficulty, 3 = much difficulty, and 4 = unable to do, was provided for each question.

Procedure for Evaluation of Psychometric Properties

Since validating an item bank of questions requires appropriate psychometric measures, we evaluated the dimensionality using factor analysis, reliability using Cronbach's alpha, and internal validity using G-IRT. Since the questions are designed to demonstrate the single dimension of activities of daily living, we deemed a single factor dimension as suitable. For the reliability analysis, alpha values greater than 0.70 were considered acceptable.

Statistical Methods

All statistical analyses were performed using Stata version 8.0 for Linux (Stata Corporation, College Station, Tex) and Multilog for Windows (SSI Scientific Software International, Lincolnwood, Ill). Initially, descriptive analyses using means and percentages with 95% confidence intervals were used to establish the demographic and clinical characteristics of the sample. Two questions independent of the item bank (1) *limitations that keep you from working* and (2) *average level of physical activity each day*, within the NHANES survey were descriptively investigated to determine the severity of functional limitations associated with a stroke. Factor structure was analyzed using factor analysis with promax rotation. Promax rotation was selected because it provides conceptually simple results, is a calculated element of varimax rotation, and forces structures toward bipolar results.³⁰ Correlation analyses examined Cronbach alpha reliability coefficients. A G-IRT model IRT analysis was performed identifying slope parameters, threshold levels, and item characteristic curves.

RESULTS

The study group consisted of 364 community-dwelling individuals who reported a history of stroke (Table 1). These survivors of a stroke (191 male, 173 female) had an average age of 71.8 ± 12.3 years (range 22-88 years). The majority of respondents were White (54.7%), but also included Blacks (21.9%), and Hispanics (21.1%). Slightly less than a fourth (24.4%) of the participants reported less than a high school education and only 12.9% indicated an education beyond a high school diploma. Nearly 52% failed to report their educational level.

Both questions used as internal measures of severity (which were independent of the item bank) demonstrated that the sample did include subjects that demonstrated ADL losses. Slightly over 43.1% of respondents indicated the effects of the stroke prevented them from working while almost 55.5%

indicated they sat during the day and did not walk about very much. Each of the 16 line items of the item bank also demonstrated wide variations in patient severity. Depending on the activity performed, 10% to 50% of respondents indicated they were 'unable to do' or 'have much difficulty' performing selected activities of daily living.

Factor analysis, with promax rotation (Table 2) of the 16 items yielded a one-factor solution indicating unidimensionality. The single construct, 'activities,' accounted for the majority (86%) of the data variance. The Cronbach alpha reliability index for the single construct was 0.95, well above the predetermined acceptance of 0.70.

Figure 1 outlines the item characteristic curves of the 16 line items of the targeted item bank questions. For each panel shown in the figure, the horizontal axis represents the latent underlying continuum of ADL impact associated with impairments associated with a stroke. This continuum is expressed in z-score units ranging from -3 to 3. Negative z-scores, or scores to the left of 0, represent lower levels of ADL impairment associated with stroke and positive z-scores (to the right of 0) represent higher levels of ADL impairment associated with stroke. The curves on each horizontal axis of each graph in Figure 1 represent variation between the Likert responses within the item. The responses are shown from left to right, with the first line curve representing the response '1 = no difficulty,' the second representing the response '2 = some difficulty,' the third representing the response '3 = much difficulty,' and the fourth representing the response of '4 = unable to do.' Six of the 16 items (Items, 1 (walking for a quarter of a mile), 2 (walking up 10 steps without resting), 3 (stooping, crouching, or kneeling), 4

(lifting or carrying something as heavy as 10 pounds), 5 (doing chores around the house like vacuuming, sweeping, dusting, or straightening up), and 11 (standing or being on your feet for about 2 hours) yielded negative values for the scores of 1 = no difficulty or 2 = some difficulty. This suggests that lower Likert values (ie, 1 = no difficulty or 2 = some difficulty) for these items may not accurately reflect minimal difficulties associated with ADL dysfunction. Respondents only minimally affected by a stroke appeared less challenged by these activities versus others. Furthermore, the responses to item 14 (going out to things like shopping, movies, or sporting events) were tightly oriented suggesting very little difference in response values between 1 = no difficulty, 2 = some difficulty, 3 = much difficulty, and 4 = unable to do.

Table 3 outlines the estimated item parameters for each question in the item bank. The vertical axis of each curve height represents the probability from 0 to 1 that the answer to the Likert type question is associated with the degree of ADL dysfunction. Slopes that rise and fall sharply are indicative of high discrimination.²⁶ An 'ideal' item characteristic curve would demonstrate equally measured peak for the outer 2 curves and symmetrically oriented middle 2 curves with equidistance measures between curve intersections.³¹ Of the 16 items, the ideal shape is best represented by item 10 (dressing yourself, including tying shoes, working zippers, and doing buttons).

Although no specific cut-off point for probability exists, higher values are considered more compelling versus lower values.³¹ Three of the 16 items did not exhibit high discrimination. Item 12 (sitting for about 2 hours), 13 (reaching up over your head), and 16 (doing things to relax at home or for leisure

Table 1. Respondent Characteristics of the (NHANES) Sample of Patients with Report of Stroke

Variable	Category	N	%
Gender	Male	191	52.5
	Female	173	47.5
Race	White	199	54.7
	Black	80	21.9
	Hispanic	77	21.1
	Other	5	1.4
	Missing	3	0.8
Education	< High School	89	24.4
	High School or (GED)	42	11.5
	> High School	44	12.9
	Missing	189	51.9%

Table 2. Factor Analysis with Promax Rotation

NHANES Survey Items	Factor Loadings (Factor 1)
1. Walking for a quarter of a mile; that is about 2 or 3 blocks?	0.799
2. Walking up 10 steps without resting?	0.789
3. Stooping, crouching, or kneeling?	0.756
4. Lifting or carrying something as heavy as 10 pounds; like a sack of potatoes or rice?	0.741
5. Doing chores around the house like vacuuming, sweeping, dusting or straightening up?	0.825
6. Preparing your own meals?	0.721
7. Walking from one room to another on the same level (same floor)?	0.788
8. Standing up from an armless straight chair?	0.785
9. Getting in or out of bed?	0.788
10. Dressing yourself, including tying shoes, working zippers, and doing buttons?	0.721
11. Standing or being on your feet for about 2 hours?	0.765
12. Sitting for about 2 hours?	0.449
13. Reaching up over your head?	0.543
14. Going out to things like shopping, movies, or sporting events?	0.894
15. Participating in social activities such as visiting friends, attending clubs or meetings, or going to parties?	0.855
16. Doing things to relax at home or for leisure such as reading, watching TV, sewing, or listening to music?	0.555

such as reading, watching TV, sewing, or listening to music) demonstrated low slopes indicating lower levels of probability that the responses are associated with ADL dysfunction from stroke.

DISCUSSION

The majority of 16 items demonstrated strong psychometric properties including internal consistency, unidimensionality, and mostly well estimated discriminative properties. This suggests that selected items analyzed in this study accurately represent the latent variables of ADLs in a community-dwelling population of survivors of a stroke and are appropriate for item bank collection. Selecting only questions that have been validated from an item bank can reduce the burden of testing for the patients and the clinicians.¹⁶ Furthermore, because most clinicians do not use scales during clinical practice,²⁹ validation of each specific item allows clinicians to use selected preferred items within the patient history without risk of invalidating the construct of ADLs.²⁶

There are both advantages and disadvantages to using the NHANES database for analysis. One advantage is that the NHANES database has been a free source for hundreds of research studies associated with the health profile of Americans.²⁸ Because the questions within this item bank are a product of the questionnaire, low cost comparisons with past studies is possible and yields transition to additional databases. Additionally, direct item to item comparisons are possible because each item is independent from the function of a scale.²⁶ Subsequently each item is comparable within the NHANES database and external to the database, among similar populations of community-dwelling stroke survivors.²⁶ Disadvantages include accuracy of patient report of current condition, co-morbidities, and lack of control over the quality and type of data selected. In a database such as the NHANES, which contains several questions, survey fatigue may affect responses as can poor understanding of an item. In addition, the availability of the items within the dataset dictates the item bank versus novel identification by the investigator.

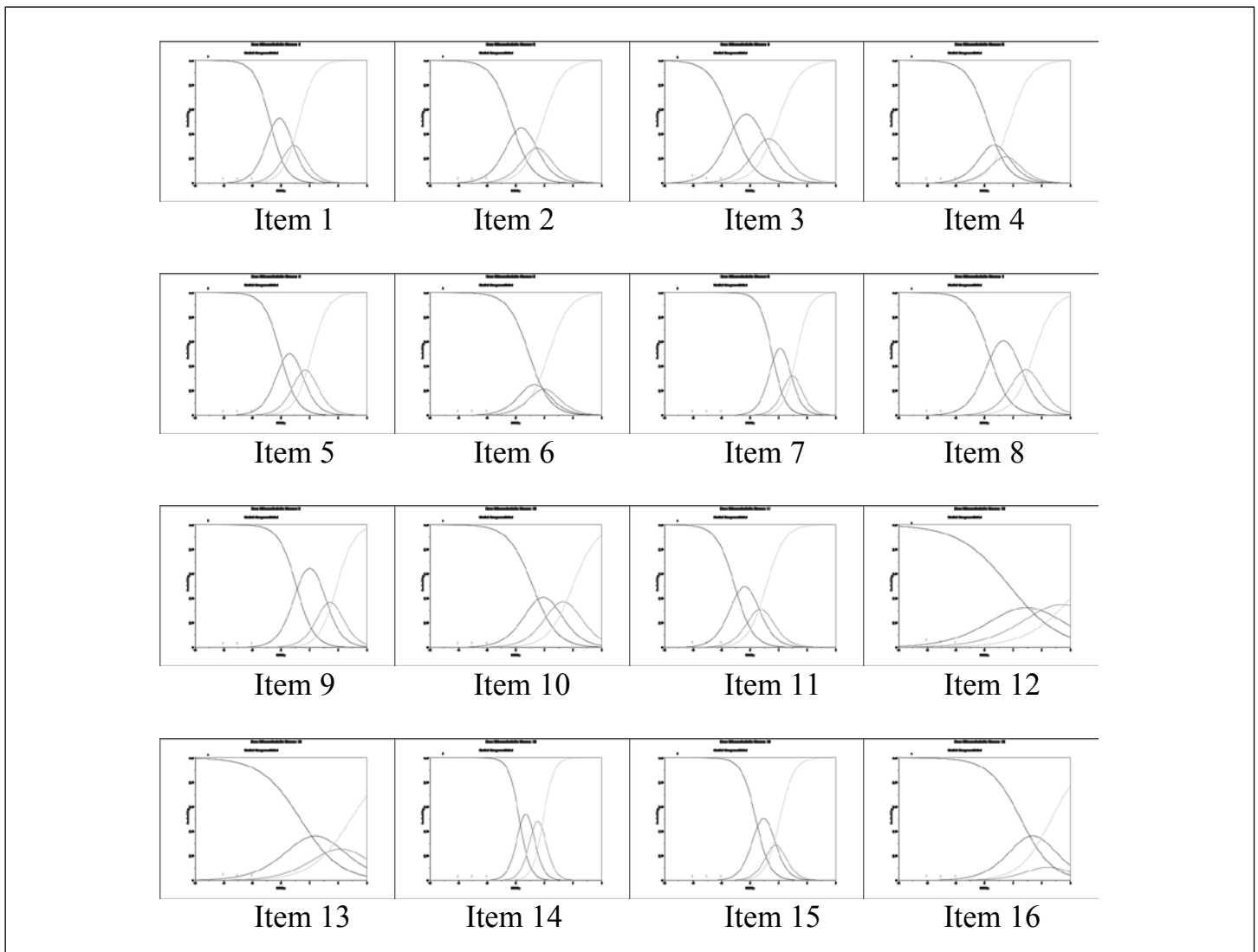


Figure 1. Item characteristic curve analysis for symptom and activity-related subscales. The vertical axis of each curve height pictorially represents the probability from 0 to 1 that the answer to the Likert type question is associated with the degree of ADL dysfunction. Flattened curves or curves that are clustered together suggest a poorly discriminant question.

Three of the 16 items did not exhibit high discrimination and should not be included within the item bank. Item 12 (*sitting for about 2 hours*), 13 (*reaching up over your head*), and 16 (*doing things to relax at home or for leisure such as reading, watching TV, sewing, or listening to music*), do not appear to be associated with ADL dysfunction in community-dwelling stroke survivors and are not considered valid questions for scale inclusion or measurement of progress. Lower Likert type selections from 6 of the items (items 1, 2, 3, 4, 5, and 11) may represent a ceiling effect and would fail to measure change in patients with low degrees of severity.

It is common for items within stroke scales to demonstrate ceiling effects.³² Line item validation through G-IRT allows the detection and adjustment of ceiling and floor effects that limit the applicability of a scale.²⁷ In some of the items analyzed

in this study, nearly 60% to 75% of respondents indicated 'no difficulty' performing activities of daily living within the constructs of the item provided. Nonetheless, these items are valuable measurement functions by rehabilitation providers and should still be considered as appropriate stroke-related questions for an item bank. It is likely these findings represent the higher level of function of the NHANES sample of community-dwelling stroke survivors and may include respondents who have completely resolved stroke related symptoms. Since the NHANES survey does not place a time limit when querying the history of a stroke, respondents may have had a long-term, chronic condition and accommodated.

Two internal measures of severity were used to identify the level of severity of the respondent group. Slightly over 43.1% of respondents indicated the effects of the stroke prevented

Table 3. Estimated Item Parameters for the Graded Response Model and the Observed Minus Expected Proportion of Responses in Each Category

Item	α (SE)	β^1 (SE)	β^2 (SE)	β^3 (SE)
1	3.61 (0.50)	-0.38 (0.08)	0.27 (0.08)	0.62 (0.10)
2	2.85 (0.39)	-0.15 (0.08)	0.53 (0.12)	0.94 (0.13)
3	2.55 (0.33)	-0.63 (0.10)	0.36 (0.10)	0.96 (0.14)
4	2.77 (0.36)	-0.11 (0.10)	0.58 (0.11)	0.89 (0.13)
5	3.46 (0.50)	-0.02 (0.08)	0.62 (0.09)	1.07 (0.12)
6	2.75 (0.45)	0.46 (0.10)	0.83 (0.12)	1.14 (0.17)
7	4.45 (0.89)	0.77 (0.08)	1.32 (0.12)	1.63 (0.17)
8	2.80 (0.36)	0.16 (0.09)	1.16 (0.15)	1.72 (0.22)
9	3.27 (0.55)	0.54 (0.08)	1.47 (0.19)	1.95 (0.28)
10	2.36 (0.39)	0.59 (0.11)	1.32 (0.18)	1.98 (0.30)
11	3.15 (0.39)	-0.54 (0.09)	0.15 (0.09)	0.56 (0.11)
12	1.14 (0.23)	0.87 (0.21)	2.06 (0.39)	3.34 (0.70)
13	1.45 (0.26)	0.68 (0.16)	1.72 (0.31)	2.44 (0.42)
14	5.43 (0.95)	0.13 (0.06)	0.57 (0.08)	0.96 (0.09)
15	4.06 (0.61)	0.20 (0.07)	0.75 (0.10)	1.04 (0.11)
16	1.83 (0.40)	1.27 (0.20)	2.10 (0.36)	2.33 (0.42)

Each β column represents the transition between adjacent Likert values. Subsequently, the selections of 1 = no difficulty to 2 = some difficulty are represented by β^1 , to 2 = some difficulty to 3 = much difficulty are represented by β^2 , and 3 = much difficulty to 4 = unable to do are represented by β^3 .

them from working while almost 55.5% indicated they sat during the day and did not walk about very much. This suggests that nearly half of the respondent pool did not have severe symptoms. Although an IRT analysis is less affected by severity of patient population than validity measures associated with classical test theory, biases toward one direction could flatten the item characteristic curves yielding less discriminatory values.²⁶

The item bank from the NHANES survey provides validated patient-based individual questions along with a number of stroke-specific scales described within the literature. Two of the most common ADL scales reported in the literature are the Barthel Index, a therapist observation-based scale, and the Frenchay Activities Index, which is based on patient interview. The Barthel Index spans 2 domains of basic ADLs (basic mobility and self-care). However it includes 2 non-ADL questions related to impairment or function.²⁴ The Frenchay Activities Index spans 4 domains of extended ADLs (household, community, social/recreational, and cognitive) including 1 non-ADL question.²⁴ The NHANES item bank provides a collection of

questions, which can be used independently of one another, and spans the entire range of ADLs in community dwelling stroke patients throughout their continuum of care. This allows clinicians to pick validated questions from the survey best suited for each patient in order to quantify their level of ADL dysfunction. Furthermore, the NHANES item bank is patient report, similar to the Frenchay Activities Index.

There are limitations to this study. Development of an item bank provides the clinician with individually validated questions for use within a given population. What is unknown from this finding is the responsiveness of each individual item and comparisons of criterion validity with other ADL scales. One limitation associated with use of secondary data is the inability to manipulate selected variables. Three of the 16 items did not accurately represent ADL dysfunction and may require manipulation. Furthermore, although participants reported a recent history of stroke, concomitant conditions were not controlled. Concomitant condition could have influenced the report of ADL dysfunction for the individual line items.

CONCLUSION

Thirteen items associated with activities of daily living challenges were validated individually in a sample of community dwelling stroke survivors. Clinicians can be confident that each of these items are scale independent, are valid for measurement of functional activities, and may be applicable for use in another scale for assessment of functional change in community-dwelling stroke survivors.

Further studies should prospectively explore the responsiveness of the item bank in a sample of community dwelling stroke survivors. In addition, this item bank should be validated in other pathologies such as cancer, low back pain, cervical pain, and arthritis, all populations well represented within the NHANES database.

REFERENCES

1. Thom T, Haase N, Rosamond W, et al. Heart disease and stroke statistics--2006 update: a report from the American Heart Association Statistics Committee and Stroke Statistics Subcommittee. *Circulation*. 2006;113:e85-151.
2. Feigin VL, Lawes CM, Bennett DA, Anderson CS. Stroke epidemiology: a review of population-based studies of incidence, prevalence, and case-fatality in the late 20th century. *Lancet Neurol*. 2003;2:43-53.
3. Centers for Disease Control and Prevention (CDC). Prevalence of disabilities and associated health conditions among adults--United States, 1999. *MMWR Morb Mortal Wkly Rep*. 2001;50:120-125.
4. Jorgensen HS, Nakayama H, Raaschou HO, Olsen TS. Stroke. Neurologic and functional recovery the Copenhagen Stroke Study. *Phys Med Rehabil Clin N Am*. 1999;10:887-906.
5. Duncan P, Studenski S, Richards L, et al. Randomized clinical trial of therapeutic exercise in subacute stroke. *Stroke*. 2003;34:2173-2180.
6. Lai SM, Studenski S, Duncan PW, Perera S. Persisting consequences of stroke measured by the Stroke Impact Scale. *Stroke*. 2002;33:1840-1844.
7. Wade DT. Evaluating outcome in stroke rehabilitation (quality control and clinical audit). *Scand J Rehabil Med Suppl*. 1992;26:97-104.
8. Wade DT. Is stroke rehabilitation worthwhile? *Curr Opin Neurol Neurosurg*. 1993;6:78-82.
9. Reker DM, Hoenig H, Zolkewitz MA, et al. The structure and structural effects of VA rehabilitation bedservice care for stroke. *J Rehabil Res Dev*. 2000;37:483-491.
10. Ottenbacher KJ, Jannell S. The results of clinical trials in stroke rehabilitation research. *Arch Neurol*. 1993;50:37-44.
11. Langhorne P, Williams BO, Gilchrist W, Howie K. Do stroke units save lives? *Lancet*. 1993;342:395-398.
12. Cook C, Stickley L, Ramey K, Knotts VJ. A variables associated with occupational and physical therapy stroke rehabilitation utilization and outcomes. *J Allied Health*. 2005;34:3-10.
13. Alexander H, Bugge C, Hagen S. What is the association between the different components of stroke rehabilitation and health outcomes? *Clin Rehabil*. 2001;15:207-215.
14. Leidy NK. Using functional status to assess treatment outcomes. *Chest*. 1994;106:1645-1646.
15. Testa MA, Simonson DC. Assessment of quality-of-life outcomes. *N Engl J Med*. 1996;334:835-840.
16. Holman R, Weisscher N, Glas CA, et al. The Academic Medical Center Linear Disability Score (ALDS) item bank: item response theory analysis in a mixed patient population. *Health Qual Life Outcomes*. 2005;3:83.
17. Bode RK, Lai JS, Cella D, Heinemann AW. Issues in the development of an item bank. *Arch Phys Med Rehabil*. 2003;84:S52-60.
18. McHorney CA. Ten recommendations for advancing patient-centered outcomes measurement for older persons. *Ann Intern Med*. 2003;139:403-409.
19. Lin JH, Wang WC, Sheu CF, Lo SK, Hsueh IP, Hsieh CL. A Rasch analysis of a self-perceived change in quality of life scale in patients with mild stroke. *Qual Life Res*. 2005;14:2259-2263.
20. van Hartingsveld F, Lucas C, Kwakkel G, Lindeboom R. Improved interpretation of stroke trial results using empirical Barthel item weights. *Stroke*. 2006;37:162-166.
21. Dallmeijer AJ, Dekker J, Roorda LD, et al. Differential item functioning of the Functional Independence Measure in higher performing neurological patients. *J Rehabil Med*. 2005;37:346-352.
22. Hsueh IP, Wang WC, Sheu CF, Hsieh CL. Rasch analysis of combining two indices to assess comprehensive ADL function in stroke patients. *Stroke*. 2004;35:721-726.
23. Tennant A, Penta M, Tesio L, et al. Assessing and adjusting for cross-cultural validity of impairment and activity limitation scales through differential item functioning within the framework of the Rasch model: the PRO-ESOR project. *Med Care*. 2004;42:137-48.
24. Lindeboom R, Vermeulen M, Holman R, De Haan RJ. Activities of daily living instruments: optimizing scales for neurologic assessments. *Neurology*. 2003;60:738-742.
25. Hays RD, Morales LS, Reise SP. Item response theory and health outcomes measurement in the 21st century. *Med Care*. 2000;38:1128-42.
26. van der Linden W, Hambleton R. *Handbook of Modern Item Response Theory*. New York, NY: Springer; 1997.
27. Pietrobon R, Coeytaux RR, Carey TS, Richardson WJ, DeVellis RF. Standard scales for measurement of functional outcome for cervical pain or dysfunction: a systematic review. *Spine*. 2002;27:515-522.
28. NHANES 1999-2000 Public Data Release File Documentation. Available at:<http://www.cdc.gov/nchs/about/major/nhanes/NHANES19992000FAQS.htm>. Accessed November 30, 2004.
29. McKeivitt C, Redfern J, La-Placa V, Wolfe CD. Defining and

- using quality of life: a survey of health care professionals. *Clin Rehabil.* 2003;17:865-870.
30. Reyment RA, Jöreskog KG. *Applied Factor Analysis in the Natural Sciences.* Cambridge, Mass: Cambridge University Press; 1993.
31. Kawata AK, Coeytaux RR, Devellis RF, Finkel AG, Mann JD, Kahn K. Psychometric properties of the HIT-6 among patients in a headache-specialty practice. *Headache.* 2005;45:638-643.
32. Liu M, Chino N, Tuji T, Masakado Y, Hase K, Kimura A. Psychometric properties of the Stroke Impairment Assessment Set (SIAS). *Neurorehabil Neural Repair.* 2002;16:339-351.

Appendix A. The Items within the Item Bank Analysis

The following questions ask about difficulties you may have doing certain activities because of a health problem. A “health problem” is any long-term physical, mental, or emotional problem or illness (not including pregnancy). Please circle the number that most closely represents your difficulty. By yourself and without using any special equipment, how much difficulty do you have?

Item	No Difficulty	Some Difficulty	Much Difficulty	Unable to Do
1. Walking for a quarter of a mile; that is about 2 or 3 blocks?	1	2	3	4
2. Walking up 10 steps without resting?	1	2	3	4
3. Stooping, crouching, or kneeling?	1	2	3	4
4. Lifting or carrying something as heavy as 10 pounds; like a sack of potatoes or rice?	1	2	3	4
5. Doing chores around the house like vacuuming, sweeping, dusting or straightening up?	1	2	3	4
6. Preparing your own meals?	1	2	3	4
7. Walking from one room to another on the same level (same floor)?	1	2	3	4
8. Standing up from an armless straight chair?	1	2	3	4
9. Getting in or out of bed?	1	2	3	4
10. Dressing yourself, including tying shoes, working zippers, and doing buttons?	1	2	3	4
11. Standing or being on your feet for about 2 hours?	1	2	3	4
12. Sitting for about 2 hours?	1	2	3	4
13. Reaching up over your head?	1	2	3	4
14. Going out to things like shopping, movies, or sporting events?	1	2	3	4
15. Participating in social activities such as visiting friends, attending clubs or meetings, or going to parties?	1	2	3	4
16. Doing things to relax at home or for leisure such as reading, watching TV, sewing, or listening to music?	1	2	3	4