Interrater Reliability of AM-PAC "6-Clicks" Basic Mobility and Daily Activity Short Forms

Diane U. Jette, Mary Stilphen, Vinoth K. Ranganathan, Sandra Passek, Frederick S. Frost, Alan M. Jette

D.U. Jette, PT, DSc, FAPTA, Physical Therapy Program, MGH Institute of Health Professions, 36 First Ave, Charlestown Navy Yard, Boston, MA 02129 (USA), and Rehabilitation and Movement Science, University of Vermont, Burlington, Vermont. Address all correspondence to Dr Jette at: djette@mghihp.edu.

M. Stilphen, PT, DPT, Rehabilitation and Sports Therapy Department, Cleveland Clinic, Cleveland, Ohio.

V.K. Ranganathan, MS, MBA, Physical Medicine and Rehabilitation Department, Cleveland Clinic.

S. Passek, PT, DPT, Physical Medicine and Rehabilitation Department, Cleveland Clinic.

F.S. Frost, MD, Physical Medicine and Rehabilitation Department, Cleveland Clinic.

A.M. Jette, PT, PhD, FAPTA, School of Public Health, Health and Disability Research Institute, Boston University Medical Campus, Boston, Massachusetts.

Post a Rapid Response to this article at:
ptjournal.apta.org

Background. The interrater reliability of 2 new inpatient functional short-form measures, Activity Measure for Post-Acute Care (AM-PAC) "6-Clicks" basic mobility and daily activity scores, has yet to be established.

Objective. The purpose of this study was to examine the interrater reliability of AM-PAC "6-Clicks" measures.

Design. A prospective observational study was conducted.

Methods. Four pairs of physical therapists rated basic mobility and 4 pairs of occupational therapists rated daily activity of patients in 1 of 4 hospital services. One therapist in a pair was the primary therapist directing the assessment while the other therapist observed. Each therapist was unaware of the other's AM-PAC "6-Clicks" scores. Reliability was assessed with intraclass correlation coefficients (ICCs), Bland-Altman plots, and weighted kappa.

Results. The ICCs for the overall reliability of basic mobility and daily activity were .849 (95% confidence interval [CI]=.784, .895) and .783 (95% CI=.696, .847), respectively. The ICCs for the reliability of each pair of raters ranged from .581 (95% CI=.260, .789) to .960 (95% CI=.897, .983) for basic mobility and .316 (95% CI=−.061, .611) to .907 (95% CI=.801, .958) for daily activity. The weighted kappa values for item agreement ranged from .492 (95% CI=.382, .601) to .712 (95% CI=.607, .816) for basic mobility and .251 (95% CI=.057, .445) to .751 (95% CI=.653, .848) for daily activity. Mean differences between raters' scores were near zero.

Limitations. Raters were from one health system. Each pair of raters assessed different patients in different services.

Conclusions. The ICCs for AM-PAC "6-Clicks" total scores were very high. Levels of agreement varied across pairs of raters, from large to nearly perfect for physical therapists and from moderate to nearly perfect for occupational therapists. Levels of agreement for individual item scores ranged from small to very large.

A relatively high proportion of patients in acute care hospitals have mobility and self-care limitations.[1] A primary focus for physical therapists and occupational therapists based in hospitals is the evaluation of patients' mobility and self-care abilities to determine the need for postacute care; this process drives hospital discharge planning.[2-4] Because discharge planning may improve the efficiency of acute care and reduce costs[5] by transitioning patients from the expensive acute care setting to a post–acute care setting, assessment of patients' mobility and self-care abilities for discharge planning by physical therapists and occupational therapists happens throughout the course of hospitalization.

A standardized measure of patients' function might be useful in the discharge planning process. However, measures of patients' function have not been widely used by therapists for patients in the acute care setting.[6] Low usage may reflect limitations in the existing measures, such as their length, ambiguity in interpreting scores, or their ineffectiveness in facilitating the prediction of an appropriate discharge destination.[7] Equally problematic may be the lack of evidence for the validity and reliability of previously reported measures.

There appear to be few studies on the reliability of any clinician-determined measures of mobility and daily activity that can be used

regardless of patients' characteristics in the acute care setting; perhaps this fact contributes to the low rates of adoption of functional measures in the acute care setting. In 1994, Shields et al[8] reported on the development of a generic measure of mobility for patients in the acute care setting that was based on the level of assistance needed for 5 activities (University of Iowa Level of Assistance Scale). Other studies have examined the reliability of functional measures applied to particular types of patients in the acute care setting. For example, Van Dillen and Roach[9] examined the reliability of a measure of function for patients with neurological conditions (Acute Care Index of Function). We also identified 2 studies on the reliability of functional measures in patients with orthopedic conditions in the acute care setting.[10,11] Recently, Wellens et al[12] reported on the reliability of the interRAI Acute Care measure, a generic, comprehensive assessment tool designed to capture the function specifically of older adult patients upon admission to the acute care setting.

Recently, physical therapists and occupational therapists at Cleveland Clinic Health System hospitals piloted new standardized functional assessment instruments in the acute care setting to assess patient mobility and self-care functioning for discharge planning and related purposes.[13,14] These tools, called "6-Clicks," are short forms created from the Activity Measure for Post-Acute Care (AM-PAC) instrument, developed by researchers at Boston University.[15] One "6-Clicks" instrument assesses basic mobility, such as walking and moving from 1 position to another; the other instrument assesses daily activities, such as dressing and toileting. The advantages of these instruments include the following: they are quickly completed; they provide discrete data

that can be entered into an electronic medical record as part of the documentation of therapists' visits; they can be completed by direct observation or estimation, through clinical judgment, of patients' capabilities; and they are derived from and scored on the same standardized metric as the AM-PAC instrument and can be used in any care setting.

Two recently published studies provided evidence for the validity of the AM-PAC "6-Clicks" tools.[13,14] These studies found that the AM-PAC "6-Clicks" tools provided useful data that could contribute to the accurate prediction of hospital discharge setting and patients' need for therapy services in the acute care setting. Although studies on measurement reliability typically precede studies on validity, to date, no studies have explored the reliability of these new measures. Determining reliability is an important step in the process of determining the usefulness of these new measures, particularly given that they may rely on therapists' observations or clinical judgments of patients' probable levels of mobility and activity. The purpose of this study was to describe the interrater reliability of the AM-PAC "6-Clicks" measures when used by physical therapists and occupational therapists.

Method
Instrument
In April 2011, as part of a broad institutional initiative to provide uniform, high-quality services, AM-PAC "6-Clicks" forms and the process for their implementation were introduced to rehabilitation staff at a 1-hour in-service meeting. Therapists were instructed to assess patients at each visit. Physical therapists used the form that assesses basic mobility function, and occupational therapists used the form that assesses daily activity function.[16] Each instrument includes 6 activities, and the

---

Available With This Article at ptjournal.apta.org

- **eFigure 1:** Bland-Altman Plot for Basic Mobility Scores
- **eFigure 2:** Bland-Altman Plot for Daily Activity Scores

Table 1.
Characteristics of Raters by Discipline

Characteristic	Occupational Therapists	Physical Therapists
Sex (female), n (%)	6 (75.0)	7 (87.5)
Entry-level degree, n (%)		
Baccalaureate degree	3 (37.5)	0
Master's degree	4 (50.0)	1 (12.5)
Doctorate	1 (12.5)	7 (87.5)
Years of professional experience, median (minimum–maximum)	4.5 (1.0–20.0)	3.3 (0.5–8.0)
Years of acute care experience, median (minimum–maximum)	3.0 (1.0–16.0)	3.3 (0.5–8.0)

therapist scores the patient's level of difficulty or assistance with the activity on a scale of 1 (unable to do or total assistance required) to 4 (no difficulty or no assistance required). The basic mobility tool includes items related to turning over in bed, sitting down on and standing up from a chair, moving from lying on the back to sitting on the side of a bed, moving to and from a bed to a chair, walking in the hospital room, and climbing 3 to 5 steps. The daily activity tool includes items related to lower body dressing, upper body dressing, toileting, grooming, eating, and bathing.

Scores for each item can be determined either by direct observation of patients performing an activity or by clinical judgment about patients' probable capabilities. The sum of scores for each item provides a raw score from 6 to 24 that can be standardized to a t score for which the mean is 50 and the standard deviation is 10.[17] The range of standardized scores for the basic mobility form is 23.55 to 61.14, and a score of 42 equates to approximately 50% impairment. The range of standardized scores for the daily activity form is 17.07 to 57.54, and a score of 37 equates to approximately 50% disability. Lower scores equate to lower levels of function.

Participants

Participants included physical therapists and occupational therapists who had been routinely using the AM-PAC "6-Clicks" forms at the 3,700-bed main campus hospital of Cleveland Clinic Health System. Therapists were asked to volunteer to participate by the department manager. Because there were more volunteers than needed, participants were selected if they could be matched by clinical service and if they were not involved in other ongoing projects. Four pairs of physical therapists and 4 pairs of occupational therapists were selected. Each pair of therapists treated patients in 1 of 4 hospital services. Table 1 shows the characteristics of the participating therapists.

Procedure

Participating therapists completed initial visits to patients in pairs (pairs of physical therapists or pairs of occupational therapists). Patients were selected on the basis of their availability when the 2 therapists in a pair were able to coordinate their time. Data collection occurred over approximately 6 weeks. Pairs of physical therapists administered the basic mobility form to patients, and pairs of occupational therapists administered the daily activity form to patients. Each pair examined between 25 and 29 patients. We

used a procedure similar to that described by Van Dillen and Roach[9] for assessing the reliability of a measure in the acute care setting. One therapist in a pair was considered the primary therapist in the direct care of the patient. This primary therapist verbally and physically directed the assessment of the patient's function and recorded it in the electronic medical record using AM-PAC "6-Clicks" scores. While the primary therapist completed the assessment, the second therapist in the pair observed and recorded AM-PAC "6-Clicks" scores. Each therapist was unaware of the other's scores, and they did not communicate verbally during the assessment. Therapists in each pair assumed the role of observer for half of the patients they rated together and the role of primary therapist for the remaining half.

Because there is no standardized or recommended order of assessment for the AM-PAC "6-Clicks" items and we wanted the scoring to be based on common clinical practice, we did not attempt to manipulate or record the order. Data were entered into a clinical database that also included patients' demographic information. The institutional review boards at Cleveland Clinic and University of Vermont approved a waiver of patients' informed consent.

Data Analysis

Analyses were run separately for basic mobility and daily activity scores. The reliability of each individual item was determined with linearly weighted kappa statistics (κ_w).[18] The reliability of the total standardized score across all therapists was assessed with a one-way random-effects model to derive the intraclass correlation coefficient (ICC [1,1]). Two-way random-effects models were used to derive the ICC (2,1) for each pair of therapists. (Models considered both raters and

patients as random.) We qualitatively classified levels of agreement using the terminology recommended by Hopkins.[19] The overall ICCs and the standard deviation of the mean for the observing and primary raters' scores were used to calculate the standard error of measurement and the minimal detectable difference (MDD) at the 90% confidence level (MDD_{90}).

We also derived Bland-Altman plots to demonstrate the degree of agreement, the limits of agreement, and the relationship between the means of raters' scores and the differences in raters' scores. This approach provided a visual display of how well the measurements of the primary and observing therapists agreed. The smaller the range between the limits of agreement and the closer the mean difference was to 0, the better the agreement.

Statistical analyses were done with IBM-SPSS Statistics version 20 (IBM Corp, Armonk, New York) and MedCalc version 13.1 (MedCalc Software, Ostend, Belgium).

Role of the Funding Source
Funding from the American Physical Therapy Association supported the data collection efforts.

Results
In total, the pairs of physical therapists examined 102 patients' basic

Table 2.
Characteristics of Patients Whose Function Was Rated by Discipline

Characteristic	Occupational Therapists	Physical Therapists
Age, y, $\overline{X}$ (SD)	67.1 (12.0)	68.1 (11.6)
Sex (female), n (%)	55 (52.4)	55 (53.9)
Type of service, n (%)		
Orthopedic	29 (27.6)	25 (24.5)
Cardiovascular	26 (24.8)	27 (26.5)
Neurological	25 (23.8)	25 (24.5)
General medical/surgical	25 (23.8)	25 (24.5)

mobility function, and the pairs of occupational therapists examined 105 patients' daily activity function. The characteristics of the patients are shown in Table 2. Mean standardized scores for the primary and observing therapists are shown in Table 3.

Basic Mobility
The ICC for the overall reliability across rater pairs was .849 (95% confidence interval [CI]=.784, .895). The standard error of measurement was 3.16, and the MDD_{90} was 7.36. The ICCs for the reliability of each individual pair of physical therapist raters ranged from .581 (95% CI=.260, .789) in the orthopedic service to .960 (95% CI=.897, .983) in the medical/surgical service (Tab. 4). The κ_w values for item agreement ranged from .492 (95% CI=.382, .601) for bed mobility to .712 (95% CI=.607, .816) for bed/chair transfer (Tab. 5). Absolute agreement ranged

from 55% to 73%. The Bland-Altman plot demonstrated a mean difference between raters' scores of .136 (95% CI=−.782, 1.054) and limits of agreement between −9.02 and 9.30 (eFig. 1, available at ptjournal.apta.org). The Spearman rho for the relationship between the means of raters' scores and the differences in raters' scores was −.030 (95% CI=−.223, .165).

Daily Activity
The ICC for the overall reliability across rater pairs was .783 (95% CI=.696, .847). The standard error of measurement was 3.46, and the MDD_{90} was 8.06. The ICCs for the reliability of each individual pair of occupational therapist raters ranged from .316 (95% CI=−.061, .611) in the orthopedic service to .907 (95% CI=.801, .958) in the neurological service (Tab. 4). The κ_w values for item agreement ranged from .251 (95% CI=.057, .445) for bathing to

Table 3.
Standardized Scores by Service and Therapist Pairs

	Physical Therapists				Occupational Therapists			
	Primary		Observing		Primary		Observing	
Service	$\overline{X}$ (SD)	Range	$\overline{X}$ (SD)	Range	$\overline{X}$ (SD)	Range	$\overline{X}$ (SD)	Range
Cardiovascular	46.89 (7.85)	34.72	46.80 (9.28)	37.59	38.84 (4.42)	15.07	38.37 (5.51)	30.23
Medical/surgical	41.35 (9.37)	33.38	40.23 (8.50)	33.38	37.04 (8.85)	37.41	35.35 (8.41)	40.47
Neurological	41.09 (8.16)	32.56	41.11 (8.24)	37.59	38.27 (6.80)	26.94	39.17 (7.84)	26.94
Orthopedic	38.70 (6.12)	23.83	40.46 (6.29)	22.73	39.46 (3.10)	13.71	39.27 (5.09)	26.94
Overall	42.10 (8.42)	37.59	42.24 (8.52)	37.59	38.43 (6.07)	37.41	38.09 (6.88)	40.47

Table 4.
Reliability by Service[a]

Service	Basic Mobility Scores				Daily Activity Scores			
			95% CI				95% CI	
	n	ICC	Lower Bound	Upper Bound	n	ICC	Lower Bound	Upper Bound
Cardiovascular	27	.758	.533	.882	26	.729	.482	.868
Medical/surgical	25	.960	.897	.983	25	.833	.655	.923
Neurological	25	.921	.830	.965	25	.907	.801	.958
Orthopedic	25	.581	.260	.789	29	.316	−.061	.611
Overall	102	.849	.784	.895	105	.783	.696	.847

[a] CI=confidence interval, ICC=intraclass correlation coefficient.

.751 (95% CI=.653, .848) for upper body dressing (Tab. 5). Absolute agreement ranged from 49% to 81%. The Bland-Altman plot demonstrated a mean difference between raters' scores of −0.354 (95% CI=−1.183, 0.474) and limits of agreement between −8.75 and 8.04 (eFig. 2, available at ptjournal.apta.org). The Spearman rho for the relationship between the means of raters' scores and the differences in raters' scores was −.125 (95% CI=−.310, .068).

Discussion

To our knowledge, this is the first report on the interrater reliability of the AM-PAC "6-Clicks" forms used to assess the function of patients in the acute care setting regardless of medical condition. The physical therapists and occupational therapists who participated as raters in the present study had been using the forms for about 2.5 years when the data were collected. When the assessment forms were introduced, staff were provided with an in-service presentation and a 2-page description of and rationale for the use of the forms. Over the subsequent 2 years, additional information regarding scoring was provided both formally and informally. Despite this modest approach to initiating the use of the forms and educating the staff, the overall ICCs for each form were very large. On the basis of the

small mean differences in raters' scores, with 95% CIs that included 0, there appeared to be no systematic differences between the raters on either form. The small correlation between the means of raters' scores and the differences in raters' scores suggested that the differences between the raters were not related to the level of patients' basic mobility function or daily activity function.

Reliability ranged from moderate to nearly perfect across various pairs of raters. For both the basic mobility form and the daily activity form, the

reliability for therapists in the orthopedic service was only moderate to large, the lowest for the pairs in the various services. This finding suggested that the reliability of the scores may have been related to the type of medical/surgical diagnosis. Because each pair of therapists rated patients in only one service, however, we were unable to determine whether the variability in reliability was related to the raters or the patients' diagnoses. It is possible that in certain conditions, for example, after surgery, patients were less likely to be asked to perform the 6

Table 5.
Item Agreement[a]

Task	Item	Weighted[b] Kappa	95% CI	
			Lower Bound	Upper Bound
Basic mobility	Bed mobility	.492	.382	.601
	Chair sit/stand	.634	.521	.747
	Moving to sit on edge of bed	.699	.601	.797
	Bed/chair transfer	.712	.607	.816
	Walk in room	.708	.608	.807
	3–5 stairs	.503	.379	.626
Daily activity	Lower body dressing	.551	.411	.692
	Bathing	.251	.057	.445
	Toileting	.338	.183	.494
	Upper body dressing	.751	.653	.848
	Personal grooming	.591	.480	.702
	Eating	.632	.497	.767

[a] CI=confidence interval.
[b] Linear weighting.

tasks because of pain, sedation, or precautions; therefore, clinical judgment was used more frequently for scoring in those conditions than for scoring in other conditions. We do not have information to support or refute this supposition, but further investigation as to how clinicians' approach to administrating AM-PAC "6-Clicks" varies by patients' characteristics may be warranted. It is also possible that a smaller range of scores for patients in the orthopedic service than for patients in the other services contributed to the relatively lower reliability coefficients for this service (Tab. 3).

Levels of agreement on the item scores varied as well, from small to very large. The 2 items with the smallest levels of agreement were bathing, including washing, rinsing, and drying, and toileting, including using a toilet, bedpan, or urinal. It is possible that the occupational therapists making the assessments did not actually observe patients' performance of these functions but scored the items using clinical judgment. Such a scenario could have contributed to the lower reliability. The item with the smallest level of agreement on the basic mobility form was bed mobility. It is possible that patients were not in bed when the physical therapists visited but were sitting in a chair. In such a scenario, the item score would have been determined by clinical judgment rather than observation of actual performance, thus affecting reliability. We have insufficient information to explore these suppositions, and the effect of clinical judgment versus direct observation on the reliability of AM-PAC "6-Clicks" measures should be explored further. It is also possible that the range of scores in our samples was restricted, potentially lowering ICCs. The standard deviations of scores reported for the samples in the present study, however, were fairly similar to those

that we reported in a previous study with considerably larger samples (basic mobility=7.3 and daily activity=5.6).[14]

Only one study has reported the reliability of a larger set of AM-PAC items.[20] The patients in that study were receiving rehabilitation services in acute care rehabilitation hospitals, transitional care units, and outpatient clinics or at home, and the items were completed on the basis of interviews using patients' self-reports. Test-retest reliability was determined on the basis of 2 interviews approximately 3 days apart. The reliability coefficients for 12 AM-PAC items measuring basic mobility function and 16 items measuring daily activity function were .97 and .96, respectively.

In addition to the fact that we examined interrater reliability rather than test-retest reliability, several important differences between the present study and that reported by Andres et al[20] may explain the variations in the findings. In the previous study, some of the items were similar to those included in the AM-PAC "6-Clicks" forms; however, additional items were part of the assessment. The AM-PAC "6-Clicks" scores were reported by the patients, whereas in the acute care setting, the scores are determined by the clinician. Additionally, clinicians using the tools in the acute care setting may determine scores on the basis of their clinical judgment or through direct observation of a patient's performance. It is likely that 2 clinicians' judgments about patients' potential for completing tasks will vary to some degree; such variation may explain the lower reliability in the present study. It is also possible that the primary therapist who physically assessed the patient in the present study would have obtained clues about the patient's need for assistance that the observing therapist

would not have obtained; this difference might have affected the agreement about scores.

We chose to examine interrater reliability during a patient encounter with one rater physically assessing the patient and the other observing because we wanted each therapist to assess the same patient at the same time. Patients with acute illness necessitating hospitalization may experience changes in medical status over short time intervals; in addition, patients are likely to fatigue easily or experience increased pain with any activity. Therefore, we did not believe that we could select a logical time between visits to patients during which their status was not likely to change.

Using the test-retest reliability estimates reported by Andres et al,[20] Jette et al[21] calculated MDD_{90}s of 4.28 and 3.70 for basic mobility and daily activity computerized adaptive testing forms, respectively, in outpatient orthopedic settings. Because the ICCs reported by Andres et al[20] were larger than those reported in the present study and the MDD is a function of the reliability of a measure, the MDD_{90}s in the present study were larger than those reported by Jette et al.[21] Additionally, the MDD_{90}s in the present study were larger than those reported in a recent study on the validity of AM-PAC "6-Clicks" forms; in that study, the MDD_{90}s were based on the Cronbach alpha, a measure of internal consistency (α=.96 for each form).[14] The MDD_{90}s in the present study were based on interrater reliability estimates that were lower than the test-retest and Cronbach alpha estimates previously identified. Because of short lengths of stay and few visits by physical therapists and occupational therapists in acute care settings, it is not clear that determining changes for patients over time is an appropriate or useful application

of AM-PAC "6-Clicks" or that MDD is a valuable measure in this setting. However, as patients transition from one setting to another and AM-PAC measures are applied, clinicians may find knowledge of the MDD_{90} useful in making decisions about patients' functional progress.

Shields et al[8] reported an ICC for interrater reliability of .88 on the University of Iowa Level of Assistance Scale across patients with orthopedic, neurological, and cardiopulmonary conditions; this value is slightly higher than the value found for basic mobility in the present study. Additionally, they reported κ_w values of .41 to .80 across the 5 individual items. In a study with the same measure for patients with total hip or knee replacements, Shields et al[11] reported κ_w values of .48 to .78. The levels of agreement that they reported were smaller for walking and "sit-to-stand" items and larger for "coming-to-sitting" and stair-climbing items than those in the present study. They hypothesized that the low level of rater agreement on the item measuring the ability to walk approximately 4.5 m (15 ft) was due to a lack of variability, that is, a low level of walking function in most patients. The University of Iowa Level of Assistance Scale is somewhat more complicated to score than the AM-PAC "6-Clicks" tools and does not appear to have been widely adopted.

Kwoh et al[10] examined the reliability of occupational therapists and physical therapists in assessing the function of patients after hip and knee arthroplasties in the acute care setting. They found κ_w values of greater than .9 for "supine-to-sitting," lower body dressing, and toilet transfers items, a κ_w value of .79 for the "sit-to-stand" item, and a κ_w value of .85 for ambulation to approximately 30 m (100 ft). The levels of agreement that they reported were larger

than those reported for similar items in the present study. In contrast to the method used for determining AM-PAC "6-Clicks" scores, Kwoh et al[10] required direct observation of patients' performance after a treatment session. Using a combination of clinical judgment and observation to determine function may reduce reliability.

The interRAI Acute Care measure includes a few items similar to those included in the AM-PAC "6-Clicks" forms; for those items, the κ_w values for the reliability of 2 raters ranged from .68 for eating to .83 for toilet transfers.[12] Using a precursor to the same instrument, Gray et al[22] reported a similar range of values, from .63 for eating to .85 for toilet use. The range of agreement for items in the present study was wider; however, there are important differences in how the measures are applied. The interRAI Acute Care measure is completed with a combination of medical chart review, direct observation, and interviews of patients, family members, and nursing staff, whereas AM-PAC "6-Clicks" scoring depends solely on what clinicians observe or judge to be patients' function. Other limitations of the interRAI Acute Care measure include its application to older adult patients, its length, and its complex algorithm for scoring. As in the present study, in the study by Wellens et al,[12] assessments of patients were conducted by one research staff member while another rater listened or observed. In neither study with the interRAI Acute Care measure were raters clinicians.

Van Dillen and Roach[9] found reliability coefficients (ICCs) of .98 to .99 for bed mobility, transfers, and mobility subscales of the Acute Care Index of Function, each of which has items similar to those of the AM-PAC "6-Clicks" basic mobility form. The nearly perfect agreement may be

related to the fact that the physical therapists involved in that study had been involved in the development of the instrument and had had 2 weeks of practice before data collection and two 1-hour group sessions to ensure uniform understanding of how to administer it. In the present study, the therapists who assessed patients in the neurological service had only slightly lower levels of agreement, with less formal training and no involvement in AM-PAC "6-Clicks" development. The Acute Care Index of Function has more items and a more complicated scoring system than the AM-PAC "6-Clicks" tools.

In summary, on the basis of somewhat limited reports of the reliability of functional measurements in the acute care setting, it is difficult to compare the reliability of AM-PAC "6-Clicks" forms with those of other measures despite similar items across instruments. Reliability is related not only to training with a particular instrument but also to types of patients, the setting in which a tool is applied, the particular activities assessed, and how activities are assessed and measured. For example, variations in an item examining walking ability might include requiring the therapist to observe the patient walking, asking the patient about walking, or estimating the ability from other information; scoring might be based on various definitions of the level of assistance required; the use of an assistive device might be part of the scoring; and distance might be part of the score.

In addition to evidence for reliability offered by the present study, several factors support the adoption of the "6-Clicks" tools in clinical practice. First, the tools have an advantage over other available instruments in terms of ease of use. Second, the tools are applicable across all types

of patients in the acute care setting. Third, the tools are useful in considering recommendations for discharge setting. Fourth, the AM-PAC tools used in post–acute care settings are scored on the same scale as the "6-Clicks" tools, allowing functional changes across settings to be determined. The "6-Clicks" tools have been used primarily by physical therapists and occupational therapists. Aggregate data have been used in Cleveland Clinic Health System hospitals to describe the mobility and function of patients receiving physical therapist and occupational therapist services and to assist in management decisions about the allocation of personnel resources so that patients receive the right rehabilitation care at the right time. The data also have served as points of focus for team members in considering the appropriate discharge setting for patients.

Limitations

The present study is limited by the fact that it involved physical therapist and occupational therapist clinicians in only one setting, treating patients in only 4 services. We do not know how well the sample of patients assessed in the present study matched the population of patients in those services during the study period. However, the patients were selected only on the basis of their availability when the 2 therapists in a pair were able to coordinate their time. Additionally, each pair of raters assessed different patients in different services. Although raters were present for the same session with the patient, the therapist who observed did not have the same level of physical and verbal interaction with the patient as the primary therapist. Thus, each therapist did not have identical information for scoring purposes. Despite this shortcoming, the overall agreement across all pairs of raters was very large. We did not test intrarater

reliability because patients' conditions in the acute care setting were too changeable to allow a logical time frame for a second rating to be completed. Intrarater reliability may be important if the same therapist assesses the same patients over a length of stay in the acute care setting. Furthermore, the findings can be applied only to physical therapists and occupational therapists. Anticipating that the tools may be useful for screening patients in terms of their need for skilled rehabilitation services, we are considering future research to examine the reliability of nursing personnel in completing each tool as part of their initial nursing assessment of patients.

Although analysis of the Bland-Altman plots revealed a low Spearman rho ($-.125$) for the relationship between the means of raters' scores and the differences in raters' scores for the daily activity form, visual examination of eFigure 2 suggested that there may have been greater variability in differences between raters at the higher range of scores than at the lower range of scores. One value also appeared to be an outlier; however, we decided to complete the analysis using all values rather than restrict them. Further examination of this pattern with large sample sizes is recommended.

In conclusion, the present study suggests that the overall interrater reliability for the AM-PAC "6-Clicks" basic mobility and daily activity forms is very large. Levels of agreement vary across pairs of raters, from large to nearly perfect for physical therapists using the basic mobility form. For occupational therapists using the daily activity form, levels of agreement ranged from moderate to nearly perfect. Pairs rating patients in the orthopedic service had lower levels of agreement than those rating patients in other services. For individual item scores, the levels of

agreement ranged from small to very large. The level of agreement on the individual basic mobility items was somewhat better than that on the daily activity items. Future researchers could undertake multifaceted Rasch analysis to examine measure reliability and rater stability.[23] Further exploration of the factors associated with the various levels of agreement across services and across items also is warranted.

References

1 Muller M, Grill E, Stier-Jarmer M, et al. Validation of the comprehensive ICF core sets for patients receiving rehabilitation interventions in the acute care setting. *J Rehabil Med*. 2011;43:92–101.

2 Jette DU, Brown R, Collette N, et al. Physical therapists' management of patients in the acute care setting: an observational study. *Phys Ther*. 2009;89:1–24.

3 Griffin S. Occupational therapy practice in acute care neurology and orthopaedics. *J Allied Health*. 2002;31:35–42.

4 Jette DU, Grover L, Keck CP. A qualitative study of clinical decision making in recommending discharge placement from the acute care setting. *Phys Ther*. 2003;83:224–236.

5 Health & Social Care Joint Unit and Change Agents Team. Discharge from hospital: pathway, process and practice. Department of Health. Available at: http://www.wales.nhs.uk/sitesplus/documents/829/DoH%20-%20Discharge%20Pathway%202003.PDF. Published 2003. Accessed December 18, 2014.

6 Jette DU, Halbert J, Iverson C, et al. Use of standardized outcome measures in physical therapist practice: perceptions and applications. *Phys Ther.* 2009;89:125–135.

7 Ekstrand E, Ringsberg KA, Pessah-Rasmussen H. The Physiotherapy Clinical Outcome Variables Scale predicts length of hospital stay, discharge destination and future home facility in the acute comprehensive stroke unit. *J Rehabil Med.* 2008;40:524–528.

8 Shields RK, Leo KC, Miller B, et al. An acute care physical therapy clinical practice database for outcomes research. *Phys Ther.* 1994;74:463–470.

9 Van Dillen LR, Roach KE. Reliability and validity of the Acute Care Index of Function for patients with neurologic impairment. *Phys Ther.* 1988;68:1098–1101.

10 Kwoh CK, Petrick MA, Munin MC. Interrater reliability for function and strength measurements in the acute care hospital after elective hip and knee arthroplasty. *Arthritis Care Res.* 1997;10:128–134.

11 Shields RK, Enloe LJ, Evans RE, et al. Reliability, validity, and responsiveness of functional tests in patients with total joint replacement. *Phys Ther.* 1995;75:169–176.

12 Wellens NI, Van Lancker A, Flamaing J, et al. Interrater reliability of the interRAI Acute Care (interRAI AC). *Arch Gerontol Geriatr.* 2012;55:165–172.

13 Jette DU, Stilphen M, Ranganathan VK, et al. AM-PAC "6-Clicks" functional assessment scores predict acute hospital discharge destination. *Phys Ther.* 2014;94:1252–1261.

14 Jette DU, Stilphen M, Ranganathan VK, et al. Validity of the AM-PAC "6-Clicks" inpatient daily activity and basic mobility short forms. *Phys Ther.* 2014;94:379–391.

15 Haley SM, Coster WJ, Andres PL, et al. Activity outcome measurement for post-acute care. *Med Care.* 2004;42:I-49–I-61.

16 AM-PAC Boston University Activity Measure for Post-Acute Care. Boston Rehabilitation Outcomes Center, Boston University School of Public Health. Available at: http://www.bu.edu/bostonroc/instruments/am-pac/shortforms/. Accessed December 18, 2014.

17 Haley SM, Coster WJ, Andres PL, et al. Score comparability of short forms and computerized adaptive testing: simulation study with the Activity Measure for Post-Acute Care. *Arch Phys Med Rehabil.* 2004;85:661–666.

18 Sim J, Wright CC. The kappa statistic in reliability studies: use, interpretation, and sample size requirements. *Phys Ther.* 2005;85:257–268.

19 Hopkins WG. A new view of statistics. Available at: http://sportsci.org/resource/stats/effectmag.html. 2002. Accessed December 18, 2014.

20 Andres PL, Haley SM, Ni PS. Is patient-reported function reliable for monitoring postacute outcomes? *Am J Phys Med Rehabil.* 2003;82:614–621.

21 Jette AM, Haley SM, Tao W, et al. Prospective evaluation of the AM-PAC-CAT in outpatient rehabilitation settings. *Phys Ther.* 2007;87:385–398.

22 Gray LC, Bernabei R, Berg K, et al. Standardizing assessment of elderly people in acute care: the interRAI Acute Care instrument. *J Am Geriatr Soc.* 2008;56:536–541.

23 Lunz ME. An example of grader consistency using the multi-facet model. *Rasch Measurement Transactions.* 2007;21:1101–1102.