An Assessment Tool Translation Study

Joan L. Buchanan, Ph.D., Patricia L. Andres, M.S., P.T., Stephen M. Haley, Ph.D., P.T., Susan M. Paddock, Ph.D., and Alan M. Zaslavsky, Ph.D.

Policymakers hoped to substitute a new, multi-purpose, functional assessment instrument, the minimum data set post-acute care (MDS-PAC), into the planned prospective payment system (PPS) for inpatient rehabilitation hospitals. PPS design requires a large database linking treatment costs with measures of the need for care, so the PPS was designed using the functional independence measure (FIM™) database linked to Medicare hospital claims. An accurate translation from the MDS-PAC items to FIM™-like items was needed to ensure payment equity under the substitution. This article describes the translation efforts and some of the problems that led policymakers to abandon the effort.

INTRODUCTION

Over the past 20 years, functional status measurement has become a regular component of national health surveys, clinical care management, and evaluation and research studies of elderly persons. As attention has shifted from acute to long-term care, policymakers and providers have become increasingly interested in including functional status measures in payment, monitoring, and outcomes management systems. Providers, payers, and consumers would all benefit from comparable measures of functional status and rehabilitation outcomes across multiple care settings to facilitate equitable payment and to monitor the quality and efficiency of care delivery. The numerous assessment tools that are currently in use, particularly those used to group patients by levels of function, were designed for use in a single setting and have limited utility across different treatment settings (Morris et al., 1990; Granger et al., 1986; Fries et al., 1994; Stineman et al., 1994; and Hittle et al., 2002).

The field lacks a standardized rigorous approach to functional content and assessment techniques. Nearly all functional assessment measurements include some form of activities of daily living (ADL), that is, the ability to perform basic tasks such as eating, dressing, grooming, transferring, walking and bathing, and often also include instrumental ADLs, (IADLs), such as shopping, telephone use, laundry, medication use, managing finances, meal preparation and housework (Katz et al., 1963; Lawton and Brody, 1969; Applegate, Blass, and Franklin, 1990; McDowell and Newell, 1996; Branch and Meyers, 1987; Teresi et al., 1997). However, which subsets of tasks, how assistance is measured, and who is providing the assessment often differ (Jette, 1994). Thus, to assess either the population being treated or the quality of care rendered either across tools within a setting, or across settings when different measures are used, requires that we be able to convert items from one measurement tool to another.

This article describes some lessons learned from a large scale effort to substitute a new multi-purpose assessment tool,

the MDS-PAC, (*Federal Register*, 2000) into a patient classification and payment system designed around another functional assessment tool, the FIM™ (Granger et al., 1986). Effective payment system design requires large amounts of data relating resource needs (functional and cognitive status) to actual resource use (costs of treatment). To this end, FIM™ data were linked to Medicare rehabilitation hospital claims enabling the design of a comprehensive patient classification system (PCS) and the calculation of a set of payment weights. Because the MDS-PAC was a new instrument, the data needed to develop a PCS and an associated set of payment weights for MDS-PAC assessments did not exist. Instead, policymakers hoped that a sufficiently accurate translation from the MDS-PAC to the FIM™ could be developed so that the MDS-PAC could be used in the payment system that had been designed around the FIM™. Without an accurate translation, payments for some types of patients would no longer reflect their resource needs and could thus lead, to access difficulties and/or payment inequities to facilities. For the substitution to be successful, ADL and cognitive items in the MDS-PAC needed to be translated into similar FIM™-like items to create FIM™-like motor and cognitive scales. Ultimately, this effort to substitute the MDS-PAC for the FIM™ was not sufficiently accurate to ensure payment equity. Industry objections to the administrative burden of the new longer assessment tool in combination with our findings led policymakers to opt to continue using the FIM™. It may be that the level of translation accuracy needed for payment is notably greater than what is needed for monitoring performance, for outcomes management, or for other purposes. Nonetheless, the challenges posed in this translation effort provide important insights on the need for standardization and definitional clarity in all aspects of functional status assessment.

BACKGROUND AND PURPOSE

PPSs provide a fixed payment per case that is adjusted for differences in patient type, but is independent of the amount of service provided. Consequently, they are believed to provide an incentive for cost containment and efficient care delivery. Inpatient rehabilitation was exempted from the Medicare PPS for acute-care hospital payment when it was introduced in 1984. Rehabilitation hospitals were exempted because research at the time demonstrated that diagnoses, the basis of the Medicare PPS, were not adequate to predict resource needs in the inpatient rehabilitation population and that measures of functional status were needed (Hosek et al., 1986). At that time, there was no agreement on what measures of functional status should be used, nor were these data routinely collected. Since then, rehabilitation professionals have developed a parsimonious 18 item measure, the FIM™ (Granger et al., 1986). Further, more than one-half of all inpatient rehabilitation providers use the FIM™ and voluntarily submit these data to a centralized repository (Hamilton et al., 1987). Stineman and colleagues (1994) used the FIM™ data to develop a PCS for medical rehabilitation, called the FIM™–function related groups (FIM™-FRGs). Building on the basic FIM™-FRG design, but using larger and more recent data sets, a RAND team refined and expanded the classification system to cover all rehabilitation hospital discharges and provided the design for a PPS for rehabilitation hospitals (Carter et al., 2002).

In the 1980s and 1990s, research in another segment of the provider community, nursing facilities, was evolving along a

separate path. In response to a 1986 Institute of Medicine Study of the quality of care in nursing homes that called for improvements in nursing home quality and more patient-centered care, researchers in this community developed a comprehensive, multi-purpose instrument, the resident assessment instrument—MDS. (Morris et al., 1990). This instrument was mandated for use in all nursing facilities and is now used for care planning, patient classification for prospective payment, and quality assurance. A MDS was also developed for home health care though the payment system for home care uses an alternative assessment tool, the standardized Outcome and Assessment Information Set (OASIS) for Home Health Care (Hittle et al., 2002).

Since the introduction of the hospital PPS, hospital length of stay has fallen dramatically while discharges to all types of PAC providers (rehabilitation hospitals, nursing facilities, and home health agencies (HHAs) have increased markedly. In an effort to control costs in the PAC area, the Balanced Budget Act of 1997 mandated the introduction of PPSs for nursing facilities, rehabilitation hospitals, and HHAs. In 1998, the nursing home PPS, which uses per diem payments and a MDS-based patient classification system, resource utilization groups, version III, went into effect (Fries et al., 1994). This was followed shortly by a PPS for home health based on the OASIS with episode-based payments. With the growth in the use of PAC came increased recognition of the considerable overlap in populations being treated in each setting. Many nursing facilities now specialize in sub-acute and rehabilitation care or have special units within them to attract these patients. Thus, policymakers called for a more integrated approach to patient assessment that would cross post-acute settings. The MDS-PAC was developed as a response to this need for integra-

tion across settings. Policymakers believed that this new tool could be substituted into the proposed inpatient rehabilitation PPS that had been designed around the FIM™. A study by Williams and colleagues (1997) concluded that MDS items could be used to predict FIM™ subscale scores with reasonable accuracy, which lent credence to the proposed plan. While we know of no plans to substitute the MDS-PAC into nursing homes or HHAs (the other post-acute settings), its adoption in rehabilitation hospitals theoretically should have enabled more direct comparisons of the populations being treated in these treatment settings because ADLs items were fairly similar in the nursing home MDS and the MDS-PAC and the MDS-PAC included IADL items thought to be important in home health.

INSTRUMENTS

The FIM™ is an 18-item measure that was constructed to evaluate and monitor functional and cognitive status in inpatient rehabilitation settings. Each of the 18 items is rated on a seven-point scale from complete dependence (1) to complete independence (7). The FIM™ is often described as having two domains, a motor score domain (13 items) and a cognitive score domain (5 items). The FIM™ motor scale was created by summing the 13 individual motor item scores and the FIM™ cognitive scale score and by summing the 5 individual cognitive items. Item scoring is actually fairly complex and although the same seven standard response categories are used for all items, scoring rules differ somewhat by item. For example, the locomotion item has an explicit distance requirement and the use of modified diets for swallowing affects scoring on the eating item. Safety and the time required to complete an activity also influence scoring.

The Uniform Data System for Medical Rehabilitation (UDS$_{MR}$) developed training materials, runs a training and certification program, routinely collects FIM™ data from participating hospitals, and provides benchmarking information back to its member facilities (Hamilton et al., 1987). As part of the FIM™ training, UDS$_{MR}$ provides a detailed training manual with decision tree-like scoring instructions for the different levels of each item. Additional training materials, called FIM™ Lessons, are also available to help therapists learn the scoring nuances.

The FIM™ is a measure of disability and burden of care. It was designed for measurement by trained clinicians, but was intended to be discipline free. All 18 items must be completed so any activity that cannot be completed is scored as level 1, total assistance. Admission scores must be completed within the first 72 hours after admission, but generally refer to performance over the past 24 hours. Scoring instructions indicate that the best available information should be used and that direct observation of subject performance is preferred. At the time of this study, roughly 60 percent of the industry voluntarily used the FIM™ and submitted their data to UDS$_{MR}$.

Several studies have looked at the validity of the FIM™. Rasch analysis (1980) was used to compare the scaled measures across impairment groups and the analysis provided support for the two fundamental constructs, the motor domain and the cognitive domain (Heinemann et al., 1993). Multi-trait scaling and factor analysis were used to evaluate the FIM™ and provided supported for the cognitive and motor domains in all 20 impairment categories (Stineman et al., 1996, 1997). Others compared FIM™ scores for individuals living at three different levels of assistance in a continuing care retirement community and found that as a measure of disability, both the cognitive and motor scores discriminated across the three care levels in ways that were consistent with differences in burden of care (Pollak et al., 1996.) Another study used factor analysis on FIM™ scores for a sample of 127 consecutive admissions to a French rehabilitation hospital and found support for considering three domains within the motor score: self-care, overall body mobility, and sphincter control (Ravaud, Delcey, and Yelnik, 1999.) Construct validity in the FIM™ was evaluated by confirming that FIM™ scores varied by age, comorbidity, discharge destination, and impairment severity for patients with stroke and spinal cord injuries (Dodds et al., 1993). For specific subgroups such as patients with multiple sclerosis, traumatic brain injury, and spinal cord injury, FIM™ scores have been validated against disease-specific instruments (Sharrack et al., 1999; Corrigan, Smith-Knapp, and Granger, 1997).

The inter-rater reliability of the FIM™ has been assessed in several studies. In an early study of 89 facilities, unweighted item-level kappa coefficients ranged from 0.53 (moderate agreement) to 0.66 (good agreement). For the subset of facilities that had passed a competence exam, scores were notably higher ranging from 0.69 (good agreement) to 0.84 (excellent agreement). Intra-class correlation coefficients (ICC) for the motor domain were 0.96 and the 0.91 for the cognitive domain (Hamilton et al., 1994). Test-retest reliability was assessed on 45 cases yielding a motor score ICC=0.9 and cognitive score ICC=0.8 (Pollak et al. 1996). Inter-rater agreement varied with kappa coefficients ranging from 0.26 (poor agreement) to 0.88 (excellent agreement); ICC ranged from 0.56 to 0.99 (Sharrack et al., 1999). Another study found that while the total reliability score was good (0.83), reliability

coefficients across individual items varied markedly from 0.02 (poor agreement) to 0.77 (very good agreement) (Segal, Ditunno, Staas, 1993).

Several studies have looked at the internal consistency of FIM™ scales. One found that the FIM™ had high overall internal consistency (Dodds et al., 1993). Another found that when viewed across 20 diverse impairment categories, the motor and cognitive subscales exceeded minimum criteria for item internal consistency in 97 percent of the tests (Stineman et al., 1996.)

The MDS-PAC is a newer and more comprehensive instrument than the FIM™. It is intended to measure comparable patients across a variety of treatment settings and to serve as a care planning tool for each of these groups. Content areas on the MDS-PAC include demographic admission history, cognitive patterns, communication/vision patterns, mood and behavior patterns, functional status, bladder/bowel management, diagnoses, medical complexities, pain status, oral/nutritional status, procedures/services used, functional prognosis, and resources for discharge.

With its origins in the nursing home MDS and building on experience with a MDS home care, the MDS-PAC differed substantially from the FIM™ in both the breadth of coverage and in its approach to assessment. The MDS-PAC was viewed as a multi-purpose information-gathering tool and data collectors were instructed to consult the patient, the patient's family, and all caregivers from all shifts during the first 3 days of the patient's hospital stay, as well as to review the chart. Another difference between the instruments was that the FIM™ often instructed scorers to use the most dependent episode, while the MDS-PAC scorers were instructed to collect data over this longer timeframe and to use a more comprehensive consultation list, but to allow one or two more dependent episodes before scoring patients to a more dependent level. The MDS-PAC is scored on an eight-point scale, but scoring is from 0 to 6 going in reverse order from independent to total assistance and allowing for the activity did not occur (score 8).

As a relatively long instrument, the MDS-PAC relies more on written instructions and multiple items for completing the form. An example of this is the treatment of physical assistance in the performance of self-care activities. In the FIM™, the amount of physical assistance provided influences the level of dependence scored. In contrast, the MDS-PAC first scores the level of self-performance and then records the amount of physical assistance received in another item. Thus, in order to use the MDS-PAC information to create FIM™ motor and cognitive scale scores, rules for combining MDS-PAC elements into each of the 18 FIM™ items were needed.

A pilot study of the time to complete the MDS-PAC in rehabilitation hospitals reported 105 minutes for the first few assessments, dropping to 85 minutes after 10 or more cases. This contrasts with 20-25 minutes to complete the FIM™. A pilot inter-rater reliability study of 171 cases found that average reliability of 315 MDS-PAC items on draft 9 was 0.78 with a range of 0.51 to 1.00 (*Federal Register*, 2000).

The MDS-PAC was developed from the MDS for nursing facility residents, which has been translated and used in 15 other countries and has undergone reliability testing in 6 countries (Hawes et al., 1997; Sgadari et al., 1997). In a multi-State evaluation of the MDS, researchers found that items in key areas of functional status (cognition, ADLs, continence, and diagnoses) had intra-class correlations of 0.7 or higher, and that 89 percent of all items had intra-class correlations of 0.4 or higher (Hawes et al., 1995). The construct validity

of the MDS cognitive, ADL, and behavior domains was examined by comparing them to the Folstein Mini-Mental Status Exam, the Dementia Rating Scale scores, and the Alzheimers Disease Patient Registry physician behavior checklist. The study concluded that the MDS data demonstrated reasonable criterion validity for research purposes (Snowden et al., 1999). A confirmatory factor analysis was used on MDS data to evaluate five domains within the MDS: cognition, ADLs, time use, social quality, depression, and problem behaviors. For cognitively intact individuals and all residents together, the domain clusters except social quality were confirmed. For individuals with serious cognitive impairment, none of the domains were confirmed (Casten et al., 1998). Construct validity in the MDS was evaluated by testing the confirmed MDS domains (ADLs, cognition, time use, depression, and problem behaviors) against established clinical research measures. That study found the majority of their hypotheses were confirmed, but the validity coefficients were modest and performance for depression and problem behaviors was not as good as for ADLs, cognition and time use (Lawton et al., 1998).

INFORMATION FOR PATIENT CLASSIFICATION IN PPS

In order to classify patients for payment in the planned PPS for inpatient rehabilitation, one needs to know (1) the rehabilitation impairment category (reason for the inpatient rehabilitation admission, e.g. stroke, traumatic brain injury, lower extremity joint replacement), (2) patient age, (3) the FIM™ motor scale score (the sum of 12[1] motor items each scored from total assistance (1) to complete independence (7) and (4) the FIM™ cognitive scale

score (the sum of the 5 cognitive items each scored from total assistance (1) to complete independence (7)). The first two items are recorded using the same format on both instruments, so we focus on the latter two elements.

DEVELOPMENT OF THE INITIAL TRANSLATION

Knowing the planned substitution of the MDS-PAC for the FIM™ in the inpatient rehabilitation PPS, policymakers took several steps to facilitate and improve translation from the MDS-PAC to the FIM™. Telephone conferences between the two instrument development teams identified potential problem translation areas, leading to both item and scoring refinements for the functional status items and to the inclusion of supplemental items.

Either as part of the original MDS-PAC development process or as a result of the telephone conferences, a number of items were changed or refined from their MDS counterparts. Item translation was challenging both because the differences in the underlying approach to scoring in the two instruments and because of the desire to retain comparability with the MDS. An example of an item refinement was the dressing item. In the MDS, this is a single item. In the FIM™, dressing is two separate items, one for dressing upper body and the other for dressing lower body. The MDS-PAC uses two dressing items to parallel the FIM™.

Scoring refinements converted the six-point MDS scores independent (0), supervision (1), limited assistance (2), extensive assistance (3), total dependence (4), activity did not occur (8) to an eight-point scale

---

[1] One of the 13 motor items, tub transfer, was dropped from the scale used in the payment system because its relationship to cost was not consistent with other scale items.

by adding a setup help only, and maximal assistance (between extensive assistance and total dependence). Labels on response levels helped to establish or maintain comparability across the instruments.

The scoring for ADL assist codes was also changed. On the MDS, the ADL assist codes are scored—for no personal assistance (0)—for one person assistance (1), and—for two or more person assistance (2). On the MDS-PAC this scoring was changed to weight bearing assistance with one limb (1), two or more person physical assistance (2) and neither code applies (0). The latter then included both persons with no assistance and those receiving one person weight bearing assistance with the torso or with more than one limb.

Supplemental items added to improve comparability with the FIM™ included new items such as distance walked, stair climbing in last 24-hours, bladder appliance support, and bowel appliance support.

Policymakers also asked the MDS-PAC developers to provide an initial item-by-item translation. For the motor items, obvious counterparts existed in the two instruments. For these items, the translation reversed the orientation of the MDS-PAC's scoring scale—independent (0) to total assistance (6)—and mapped it into the corresponding FIM™ numerical values from total assistance (1) to complete independent (7) (e.g. MDS-PAC 0 became FIM™ 7). Further, it was generally agreed that the scoring level 8 activity did not occur, was used when individuals were unable to perform a task, so the MDS-PAC 8 was rescored to a FIM™ total assistance. The physical assistance codes were only used in the translation for cases where the MDS-PAC score was scored maximal assistance (5) and (1) two or more persons were needed for physical assistance. These cases were rescored to 1—total assistance (1) on the FIM™ scale.

For the five FIM™ cognitive items (comprehension, expression, social interaction, problem solving, and memory), there were no analogous individual items. The MDS-PAC contained a cognitive section with four items (comatose, memory/recall ability with four subcomponents, cognitive skills for daily decisionmaking, and indicators of delirium) and a communication section with six items (hearing, models of communication, making self understood, speech clarity, ability to understand others, and vision). A fairly complex multi-item, empirically derived translation that used both these and other items was provided by the MDS-PAC development team. This cognitive translation was used throughout the evaluation study.

EVALUATION STUDY DESIGN

Fifty FIM™-certified rehabilitation facilities (out of 180 volunteering hospitals), representing rehabilitation hospitals and units throughout the country, participated in the study. These facilities were purposively sampled to represent regions, size, rural-urban, unit-freestanding, and clustered geographically to facilitate training and data collection. Participating facilities ranged in size from 13 to 150 beds. Sixteen percent were rural and 28 percent were freestanding facilities. All facilities were previously certified in FIM™ and were participants in the UDS$_{MR}$ system. Data collectors were teams of one to four clinicians (physical and occupational therapists, nurses, speech language pathologists, etc.) from each site who attended a 2-day MDS-PAC training session and successfully completed a certification exam before data collection began. Each facility was asked to complete both the FIM™ and the MDS-PAC on all new Medicare admissions with stays beyond 3 days for an 8-week period.

This resulted in over 3,200 FIM™ and MDS-PAC pairs. One or more of three highly trained calibration teams visited each participating hospitals and rescored both the FIM™ and the MDS-PAC for three-eight current cases. Thus for approximately 200 cases we had two FIM™ and two MDS-PAC ratings.

Because we needed to train more than 250 rehabilitation professionals from across the country in a 2-week period, we used a train the trainers model. All study trainers were trained and certified on the MDS-PAC by the MDS-PAC development team trainers. Study trainers were rehabilitation professionals who were also FIM™ instructors and most had participated in pilot projects on the MDS-PAC. Thus, this was their second training session on the MDS-PAC and each had completed approximately 30 MDS-PAC cases in the earlier studies. All intended to become MDS-PAC instructors when the new instrument became official. The functional assessment portion of the trainers' training included scoring videos and written case studies and a visit to a rehabilitation facility for onsite scoring of actual patients. Each of these activities was followed by a debriefing session and discussion of the rationale for the case scoring.

In nursing homes, the MDS is completed by the nurses. However, in rehabilitation hospitals, the FIM™ was completed by rehabilitation professionals, sometimes a single individual, but often an interdisciplinary team with physical therapists completing the mobility items, occupational therapists completing the self-care items, rehabilitation nurses completing the bowel and bladder items, and speech language pathologists completing the communication and language items. We did not want to assume that nurses (the MDS data collection model) would replace the rehabilitation specialists so we asked hospitals to send four-person data collection teams for training. The study data collection teams, each with one-four members[2], included practicing rehabilitation professionals (physical and occupational therapists, nurses, speech language pathologists, recreation therapists) from each of the 50 hospitals in the study. Each team attended the 2-day training, completed a post-training assignment, and went through a telephone certification process conducted by the full-time study team field coordinator. The functional assessment training included scoring videos and written case studies each followed by a debriefing to discuss the rationale for each score.

The field coordinator maintained regular contact with the study hospitals. In addition, an 800 telephone number was provided so the scoring teams could call in questions to the field coordinator, who was supported by the MDS-PAC and FIM™ development team trainers. A document of frequently asked questions and answers was maintained on a study Web page and periodically distributed to data collectors. Regular newsletters with information on study progress, procedural updates, and scoring clarifications were sent to all data collectors.

The three calibration teams spent a month training in Boston. During the first week, they were trained and certified on the MDS-PAC by the MDS-PAC development staff trainers. They were also retrained and re-certified on the FIM™ by the FIM™ development group trainers. The functional assessment portion of their training mimicked that of the trainers with scoring videos, written case studies, and a site-visit to score actual patients each followed by a debriefing session. Calibration

_____

[2] Each hospital was asked to send one or more four-person teams to train, including a physical and an occupational therapist, a nurse, a speech language pathologist or other rehabilitation provider. Some hospitals sent only three-person teams. Actual data collection teams varied from one to four people.

team members spent the next 3 weeks working together in different cross-disciplinary combinations and rotating through four rehabilitation hospitals in the greater Boston area practicing both the MDS-PAC and the FIM™. These rotations were used to standardize scoring across all calibration team members and also provided necessary experience entering unfamiliar institutions and establishing procedures for the assessments. Final team assignments were made near the end of the training.

Refining the Translation

The first study task was to compare the actual FIM™ motor scale and item scores with those obtained from the MDS-PAC translations and summated scales. The mean FIM™ cognitive scale score was quite close to the mean PAC translation, 28.50 compared to 28.51. However, the mean FIM™ motor scale score differed from the mean PAC motor scale translation by nearly 5 points, 45.46 compared to 50.26. Individual motor items with the largest mean scoring differences were the locomotion item with a mean difference of more than 1.5 points on the seven-point scale, and grooming and toileting with mean differences of more than 0.5 points.

The evaluation team undertook a process to review and refine the translation. We began with a review of the instruments' scoring sheet instructions and the scoring manuals for both the FIM™ and the MDS-PAC. We benefited from our participation in the training courses, the certification process, and a review of the frequently asked questions. These all highlighted areas where the approach to scoring differed in the two instruments. After discrepant areas and possible refinements were identified, they were tested empirically to confirm that they led to better scoring

agreement. Scoring agreement was measured with Pearson correlations[3], and weighted and unweighted kappa statistics. Occasionally, scoring changes improved item level agreement on kappas and correlations, but not on item level means. As long as these also improved scale level mean comparisons, they were retained.

There were both fundamental and item-specific differences between the two assessment tools that we knew could not be overcome in any translation. First, the reference timeframe in the MDS-PAC is a 3-day lookback conducted on day 4, but the FIM™ scoring takes place anytime in the first 72 hours and references only the last 24 hours. Second, the FIM™ generally directs assessors to score the most dependent episode during this 24-hour period, while the PAC assessors are instructed to allow one or two more dependent episodes without scoring the more dependent level. Third, the MDS-PAC definition of total assistance (full staff performance of activity during entire period) was much more restrictive that the FIM™ definition (patient performs less than 25 percent of the effort). Fourth, the MDS-PAC included transfers on and off the bedpan as part of the toilet transfer item, but the FIM™ does not. (Buchanan et al. [2003] provide more complete listing and discussion.)

Realigning the Seven Scoring Levels

The study team review found that the seven scoring levels of the FIM™ and the translated MDS-PAC did not align properly (Table 1). The FIM™ scoring levels differentiated between complete independence (7) and modified independence (6). A FIM™ level 7 indicated that the activity was per-

---

[3] Pearson correlations quantify the linear association between two measures and thus indicate how accurately one measure can be predicted from another. They are not, however, true measures of agreement as two items or scales can be perfectly correlated, but have few, if any agreed values.

Table 1

Revised Scoring Correspondence for Converting the MDS-PAC Self-Performance Motor Items to FIM™ Like Scores

MDS-PAC Item Scoring Level	FIM™ Item Scoring Level
0 Independent	7 Complete Independence (Timely, Safe) 6 Modified Independence (With Device)
1 Set Up Help Only 2 Supervision	5 Supervision
3 Minimal Assistance (Limited Assistance)	4 Minimal Assistance
4 Moderate Assistance (Extensive Assistance)	3 Moderate Assistance
5 Maximal Assistance	2 Maximum Assistance
6 Total Assistance (Total Dependence) 8 Activity Did Not Occur	1 Total Assistance

NOTES: MDS-PAC is minimum data-set post-acute care. FIM™ is functional independence measure.

SOURCES: Buchanan, J. L. and Zaslavsky, A. M., Harvard Medical School; Andres, P. L., and Haley, S. M., Boston University; and Paddock, S. M., RAND.

formed safely and completely independent and without assistive devices. Modified independence (level 6) was used when there were safety concerns, or the patient required extra time (three times normal), or the patient used an assistive device in order to perform the activity completely independent. The MDS-PAC motor items have a single score for independence (0), regardless of the equipment used or the manner in which the activity was performed. For some motor items, the MDS-PAC devices/aids items could be used to determine if an item scored as independent should be scored as modified independence. For example, if the patient used adaptive eating utensils and was scored independent (0), then the revised translation converted the MDS-PAC score to modified independence (6) on the FIM™ scale.

When the device item was not sufficient to separate cases that should be scored as modified independence from those that are truly independent, then both groups were scored at the most likely FIM™ level (6 or 7) at admission based on the current sample and confirmed against historical FIM™ data. Thus, if in the FIM™ more cases were scored as modified independence (6) than as complete independence (7) at admission, then we revised the translation to rescore the group as a 6.

Our review also found that the FIM™ included both set up and supervision in the same score while the MDS-PAC used different scores for set up (1) and supervision (2). Thus, in the revised translation, MDS-PAC scores of 1 and 2 were mapped to a FIM™ score 5 (Table 1).

Incorporating ADL Assist Codes

After realigning the scoring categories, our review also concluded that the use of physical assistance had not been fully captured in the original translation. The FIM™ incorporates the use of physical assistance into the actual item scoring levels. In contrast, the MDS-PAC scores self-performance of an activity and the use of physical assistance in the activity separately. The ADL assist code section, particularly the one limb assist code, does not correspond precisely to the FIM™ assistance concepts and is therefore, much more difficult to incorporate into the translation. Since one limb assist is weight-bearing assistance, the translation adopted the rule that the maximum FIM™ score a patient could have with an ADL assist code of 1

was a FIM™ 4, minimum assistance. Thus, if a functional status item was scored 0, 1, or 2 which would translate to FIM™ 7, 6, or 5, but the ADL assist code was 1, then the revised translation rescored the item to a FIM™ 4. For more dependent scores, PAC 3-6, an ADL assist code of 1 did not affect the scoring.

Because the FIM™ is a burden of care instrument, any activity needing the assistance of two persons is always scored 1 (total assistance). The revised translation rescored any item where two person physical assistance was used, regardless of the PAC self-performance score, to total assistance.

Item-Specific Translation Revisions

As previously noted the locomotion item had the largest mean discrepancy, 1.5 points between the mean FIM™ score and the mean MDS-PAC score, so we carefully reviewed the scoring instructions on this item. FIM™ scoring rules instruct raters to score patients for locomotion performance at admission, by using the mode (walk or wheelchair) expected to be used most frequently at discharge. It requires that the patient move at least 150 feet for nearly all ratings above maximum assistance. The MDS-PAC scores locomotion using the most common mode at admission and has no distance criterion. We compared the MDS-PAC locomotion with the walk in facility item and found that the latter had substantially better agreement with FIM™ scores and could be used in combination with the distance item. However, by using the walk in facility item in the revised translation, we could not differentiate those expected to use a wheelchair at discharge from ambulators in the translation. The former comprised less than 15 percent of the patient population.

FIM™ scoring rules for the bladder and bowel management items are quite complex. Raters are asked to consider how they would score the level of assistance with bladder (bowel) management on the standard 1-7 scale and then to also consider how they would score the frequency of accidents on the same scale, but they record only the minimum of the two scores. The MDS-PAC on the other hand used three items, but with differing timeframes for bladder and three for bowel and records the scoring for each. The original translation reversed the orientation to FIM™ scoring order and took the minimum of the two scores. Our revised translation incorporated the information on the specific appliances being used to aid in rescoring.

Another problem was the MDS-PAC scoring for medication use in the bladder and bowel appliance support section. When a nurse passes the medication to the patient, PAC scoring instructions were to score this as maximal assistance. Since medications are routinely controlled by nursing staff in the acute rehabilitation setting, regardless of the patient's ability to participate in this activity, all patients receiving medications got scored as maximal assistance. Further, during training, assessors were directed to treat fiber supplements, which virtually all patients receive, as medications. The revised translation attempted to rescore these cases.

Evaluating the Translation

We used factor analysis on the combined set of motor items from both the FIM™ and the MDS-PAC. We repeated the factor analysis, first replacing the motor items from the MDS-PAC with their translation counterparts and then again with the revised translation items. We did this to assess whether the translation and its revi-

Table 2
FIM™ and PAC-FIM™ Mean Scores, by Type of Assessment Team[1,2,3]

| | Assessment Teams | | | |
| | Institutional | | Calibration | |
Item	FIM™ Mean Scores	PAC-FIM™ Mean Scores	FIM™ Mean Scores	PAC-FIM™ Mean Scores
Motor				
Eating	5.51	5.54	5.73	***5.77
Grooming	4.73	***4.88	4.61	4.56
Dressing-Upper Body	3.24	*3.30	3.27	***3.03
Dressing-Lower Body	4.25	***4.35	3.92	3.95
Toileting	2.99	***3.21	2.80	2.77
Bladder Management	3.37	***3.71	3.66	**3.49
Bowel Management	4.29	4.27	4.61	***4.15
Transfer-Bed/Chair	4.70	***5.20	5.33	*5.30
Transfer-Toilet	3.58	***3.70	3.56	***3.32
Transfer-Tub/Shower	3.28	***3.67	3.60	**3.44
Locomotion-Walk/Wheelchair	1.96	1.98	1.86	**2.05
Stairs	2.22	2.20	2.51	2.38
Motor Scale	45.46	***47.82	46.80	***45.76
Cognitive				
Comprehension	5.87	***5.93	5.88	5.86
Expression	5.97	5.99	5.93	5.87
Social Interaction	5.91	***5.63	6.04	***5.54
Problem Solving	5.32	5.34	5.21	5.26
Memory	5.37	***5.56	5.34	5.44
Cognitive Scale	28.50	28.51	28.53	***28.07

***$p<=0.001$.

 **$p<=0.01$.

 *$p<=0.05$.

[1] Greater score equals greater level of independence.

[2] Uses revised motor item translation and the original cognitive item translation.

[3] Statistically significant difference between the FIM™ and the MDS-PAC translated mean scores.

NOTES: FIM™ is functional independence measure. PAC-FIM™ is minimum data set-post acute care translated into FIM™-like items.

SOURCES: Buchanan, J. L. and Zaslavsky, A. M., Harvard Medical School; Andres, P. L., and Haley, S. M., Boston University; and Paddock, S. M., RAND.

sion improved the conceptual agreement between analogous items in the two instruments. We found that neither the raw items nor those from the original translation all loaded onto the same factors as the corresponding FIM™ items, while items from the revised translation did.

The revised translation reduced the mean difference in motor scores between the FIM™ and the MDS-PAC by 50 percent from the original translation (Table 2). Despite the improvement, we found that the agreement between the instruments for institutionally-based scoring teams was only moderate and absolute agreement was worse. However, when the calibration teams scored patients using both instruments, we found notably higher levels of agreement.

We used regression analysis to analyze scoring differences and found that after controlling for administrative factors and patient, and hospital characteristics, that a random effect for hospitals was significant. This implies that scoring differences varied by hospital and this variation was not explained by any of the independent variables. The effect was substantial enough to be of concern for the comparability of scoring procedures across facilities and suggests that more training is needed to adequately standardize the assessment process.

The ultimate test to determine how well this instrument substitution would work mapped each case into its payment cell, first using the FIM™ motor and cognitive scale scores and then using the translated

MDS-PAC motor and cognitive scale scores. We tried several different empirical adjustments to improve the match between the mappings. Under all of these adjustments, the level of classification agreement was low and clearly not adequate for payment purposes. Further, we found that a substantial proportion of the facilities would experience potentially important shifts in revenue. As a consequence, policymakers opted to retain the FIM™.

DISCUSSION

This translation effort failed to achieve sufficient accuracy for use in the planned payment system despite several important advantages: (1) the issue of translation from the MDS-PAC to FIM™ scales was addressed before finalizing the instrument, (2) formal communication aimed at facilitating the translation was established between the two instrument development groups, (3) the study team and its trainers benefited from training from the instrument development group trainers, and (4) data collection teams and the study team had multidisciplinary input from rehabilitation specialists. While it may be that translations needed for applications such as quality monitoring and outcomes management will not require such stringent levels of accuracy, these findings suggest we should be cautious regarding our ability to make such substitutions.

Much of the prior research on instrument performance has been in research settings or undertaken by research staff. This study aimed to test what might happen under national clinical implementation. Clinicians practicing in their regular environment performed one set of assessments and centrally trained calibration teams visited institutions for repeat assessments.

Both instruments had strengths and limitations, and their differences often provided important insights. The FIM™ was concise, which facilitated its widespread voluntary adoption and allowed a greater focus on standardization. Rehabilitation professionals told us that FIM™ language was routinely used to describe patients as professionals communicated with one another throughout the field. Documentation included both a training manual and several sets of FIM™ lessons, training tools that were designed to illustrate scoring rules.

The major limitation of the FIM™ was its scoring complexity, which was masked by deceptively simple data collection forms. The scoring rules often required implicit integration of several concepts with no way of diagnosing what aspects of the integration created scoring difficulties. The locomotion item illustrates this complexity. Assessors must score patients at admission using the mode of locomotion that is expected at discharge. Further, distance traveled and the amount of physical assistance must be incorporated into the scoring levels. Similarly, in the eating item, when swallowing problems requiring diet modifications are present, patients who are otherwise independent in eating should be scored as modified independent. While the FIM™ scoring rules were clearly stated in the training manuals and well illustrated in the FIM™ lessons, the proprietary nature of these documents precluded open dissemination and discussion. This probably contributed to our findings of scoring practices that differed from one institution to the next even after controlling for observable patient and facility characteristics.

The MDS-PAC form contained more explicit instructions for scoring each item and tended to use multiple items to build complex constructs. While this made us combine the items to achieve FIM™-like scoring, it enabled the study team to diag-

nose some problem areas. For example, when we looked at scoring reliability for the translation, we were able to look at the reliabilities of the translated item and of each of its component parts. For the functional status items, the physical assistance items had much lower scoring reliabilities than the self-performance items.

Having the explicit components in the MDS-PAC also enabled us to determine that FIM™ scorers did not always incorporate concepts such as distance and modified diets into their scoring. This led us to recommend that these items be added to the officially adopted assessment tool to remind scorers to consider them. At worst, their inclusion allows us to explicitly identify this as a scoring problem with potential for correction.

Other strengths of the MDS-PAC include: (1) its greater breadth of coverage of substantive areas, (2) its potential for use in care planning and monitoring quality of care, and (3) its potential for application across multiple settings.

Its major limitations are its length and its lack of standardization in response formats and timeframes. Much of the training session was devoted to the mechanics of the form, that is, where to put checks versus numerical scores and where blanks were allowed, leaving inadequate time to address definitional clarity and standardization in assessment. For example, the rehabilitation professionals clearly wanted additional guidance on selecting functional status performance levels (especially distinguishing minimal, moderate, maximum, and total assistance). They also wanted clearer definitions for terms like partial versus full loss of voluntary motor control. An example of the lack of standardization on timeframe is in the bowel management section. Bowel continence is coded over the last 7-14 days even though on the admission assessment, bowel appliances are coded over the last 3 days and bowel appliance support is coded over the last 24 hours.

In conclusion, this evaluation clearly points to the need for a unified common conceptual framework and a rigorous standardized approach to the content of functional assessment measures and to the assessment techniques used. Standardization may be facilitated when the item set is relatively small and scoring rules are simplified or when complex scoring rules can be broken down into simpler, explicit steps. These findings should be kept in mind as we move forward with the development, refinement, and evaluation of the International Classification of Functioning.

REFERENCES

Applegate, W.B., Blass, J.P., and Williams F.T.: Instruments for the Functional Assessment of Older Patients. *New England Journal of Medicine* 322(17):1207-1214, 1990.

Branch, L.G., and Meyers, A.R.: Assessing Physical Function in the Elderly. *Clinics in Geriatric Medicine* 3(1):29-51, 1987.

Buchanan, J.L., Andres, P. L., Haley, S.M., et al.: *Final Report on the Instruments for PPS*. MR-1501-CMS. RAND. Santa Monica, CA. 2002.

Carter, G.M., Buntin, M.B., Hayden, O., et al.: *Analyses for the Implementation of Medicare's Inpatient Rehabilitation Prospective Payment System*. MR-1501-CMS. RAND. Santa Monica, CA. 2002.

Casten, R., Lawton, M.P., Parmelee, P.A., and Kleban, M.H.: Psychometric Characteristics of the Minimum Data Set I: Confirmatory Factor Analysis. *Journal of the American Geriatric Society* 46(6):726-735, June 1998.

Corrigan, J.D., Smith-Knapp, K., and Grange, C.V.: Validity of the Functional Independence Measure for Persons With Traumatic Brain Injury. *Archives of Physical Medicine and Rehabilitation* 78(8):828-834, August 1997.

Dodds, T.A., Martin, D.P., Stolov, W.C., and Deyo, R. A.: A Validation of the Functional Independence Measurement and Its Performance Among Rehabilitation Inpatients. *Arch Phys Med Rehabilitation* 74(5):531-536, May 1993.

Federal Register: Medicare Program: Prospective Payment System for Inpatient Rehabilitation Facilities; Proposed Rule. 65 FR 26281, May 5, 2000.

Fries, B.E., Schneider, D.P., Foley, W.J., et al.: Refining a Case-Mix Measure for Nursing Homes: Resource Utilization Groups (RUG III). *Medical Care* 32(7):668-685, July 1994.

Granger, C.V., Hamilton, B.B., Keith, R.A., et al.: Advances in Functional Assessment for Medical Rehabilitation. In Lewis, C.B. (ed.): *Topics in Geriatric Rehabilitation* 1:59-74, 1986.

Hamilton, B.B., Granger, C.V., Sherwin, F.S., et al.: A Uniform National Data System for Medical Rehabilitation. In Fuhrer, M.J. (ed.): *Rehabilitation Outcomes: Analysis and Measurement*. Baltimore, MD. Brookes Publishing Co. 1987.

Hamilton, B.B., Laughlin, J.A., Fiedler, R.C., and Granger, C.V. Inter-rater Reliability of the Seven Level Functional Independence Measure (FIM™) *Scandinavian Journal of Rehabilitation Medicine* 26:115-119, 1994.

Hawes, C., Morris, J. N., Phillips, C. D., et al.: Development of the Nursing Home Resident Assessment Instrument in the U.S.A. *Age and Aging* 26(Suppl 2):19-25, September 1997.

Hawes, C., Morris, J. N., Phillips, C. D., et al.: Reliability Estimates for the Minimum Data Sets for Nursing Home Resident Assessment and Care Screening (MDS). *The Gerontologist* 35(2):172-178, April 1995.

Heinemann, A.W., Linacre, J.M., Wright, B.D., et al.: Relationships Between Impairment and Physical Disability as Measured by the Functional Independence Measure. *Archives of Physical Medicine and Rehabilitation* 74(6):566-573, June 1993.

Hittle, D. F., Crisler, K. S., Beaudry, J. M. et al.: *OASIS and Outcome-Based Quality Improvement in Home Health Care: Research and Demonstration Findings, Policy Implications, and Considerations for Future Change*. University of Colorado Health Sciences Center. March 2002. Internet address: www.cms.hhs.gov/providers/hha.(Accessed 2003.)

Hosek, S., Kane, R., Carney, M., et al.: *Charges and Outcomes for Rehabilitative Care: Implications for the Prospective Payment System*. R-3424-HCFA. RAND. Santa Monica, CA. 1986.

Institute of Medicine Study: *Improving the Quality of Care in Nursing Homes*. National Academy of Sciences Press. Washington, DC. 1986.

Jette, A.M.: How Measurement Techniques Influence Estimates of Disability in Older Populations. *Social Science and Medicine* 38(7):937-942, July 1994.

Katz, S., Ford, A.B., Moskowitz, R.W., et al.: Studies of Illness in the Aged. The Index of the ADL: A Standardized Measure of Biological and Psychological Functioning. *Journal of the American Medical Association* 185(12):914-919, September 1963.

Lawton, M.P., and Brody, E.M.: Assessment of Older People: Self-Maintaining and Instrumental Activities of Daily Living. *Gerontologist* 9(3):179-186, September 1969.

Lawton, M.P., Casten, R., Parmelee, P.A., et al.: Psychometric Characteristics of the Minimum Data Set II: Validity. *Journal of the American Geriatric Society* 46(6):736-744, June 1998.

McDowell, I. and Newell, C.: *Measuring Health: A Guide to Rating Scales and Questionnaires*. Oxford University Press, New York, NY. 1996.

Morris, J.N., Hawes, C., Fries, B.E., et al.: Designing the National Resident Assessment Instrument for Nursing Homes. The *Gerontologist* 30(3):293-307, June 1990.

Pollak, N., Rheault, W., and Stoecker, J.L.: Reliability and Validity of the FIM™ for Persons Aged 80 Years and Above From a Multilevel Continuing Care Retirement Community. *Archives of Physical Medicine and Rehabilitation* 77(10):1056-1061, October 1996.

Rasch, G.: *Probabilistic Models for Some Intelligence and Attainment*. University of Chicago Press. Chicago, IL. 1980.

Ravaud, J.F., Delcey, M., and Yelnik, A.: Construct Validity of the Functional Independence Measure (FIM™): Questioning the Unidimensionality of the Scale and the Value of FIM™ Scores. *Scandinavian Journal of Rehabilitation Medicine* 31(1):31-41, March 1999.

Segal, M.D., Ditunno, J.F., and Staas, W.E.: Interinstitutional Agreement of Individual Functional Independence Measure (FIM™) Items Measured at Two Sites on One Sample of SCI Patients. *Paraplegia* 31:622-631, 1993.

Sgadari, A., Morris, J.N., Fries, B.E., et al.: Efforts to Establish the Reliability of the Resident Assessment Instrument. *Age and Aging* 26(Suppl 2):27-30, September 1997.

Sharrack, B., Hughes, R.A.C., Soudain, S., and Dunn, G.: The Psychometric Properties of Clinical Rating Scales Used in Multiple Sclerosis. *Brain* 122(1):141-159, January 1999.

Snowden, M., McCormick, W., Russo, J., et al.: Validity and Responsiveness of the Minimum Data Set. *Journal of the American Geriatric Society* 47(8): 1000-1004, August 1999.

Stineman, M.G., Goin, J.E., Hamilton, B.B., and Granger, C.V.: A Case Mix Classification System for Medical Rehabilitation. *Medical Care* 32(4):366-79, April 1994.

Stineman, M.G., Shea, J.A., Jette, A., et al.: The Functional Independence Measure: Tests of Scaling Assumptions, Structure, and Reliability Across 20 Diverse Impairment Categories. *Archives of Physical Medicine and Rehabilitation* 77(11):1101-1108, November 1996.

Stineman, M.G., Jette, A.J., Fiedler, R.C., and Granger, C.V.: Impairment-Specific Dimensions Within the Functional Independence Measure. *Archives of Physical Medicine and Rehabilitation* 78(6):636-643, June 1997.

Teresi, J.A., Lawton, M.P., Holmes, D., and Ory, M. (eds.): *Measurement in Elderly Chronic Care Populations.* Springer Publishing, New York, NY. 1997.

Williams, B.C., Li, Y., Fries, B.E., and Warren, R.L.: Predicting Patient Sores Between the Functional Independence Measure and the Minimum Data Set: Development and Performance of a FIM™-MDS "Crosswalk". *Archives of Physical Medicine and Rehabilitation* 78(1):48-54, January, 1997.