
Measuring Function for Medicare Inpatient Rehabilitation Payment

Grace M. Carter, Ph.D., Daniel A. Relles, Ph.D., Gregory K. Ridgeway, Ph.D., and Carolyn M. Rimes, M.A.

We studied 186,766 Medicare discharges to the community in 1999 from 694 inpatient rehabilitation facilities (IRF). Statistical models were used to examine the relationship of functional items and scales to accounting cost within impairment categories. For most items, more independence leads to lower costs. However, two items are not associated with cost in the expected way. The probable causes of these anomalies are discussed along with implications for payment policy. We present the rules used to construct administratively simple, homogeneous, resource use groups that provide reasonable incentives for access and quality care and that determine payments under the new IRF prospective payment system (PPS).

INTRODUCTION

The ability of patients to perform various functions is currently recorded in administrative data related to several types of health services in Canada and the U. S. Functional status information is used for care planning, to measure quality of care, and to adjust payments for case mix under various Medicare PPSs. The functional independence measure (FIM™) has historically been used for care planning and quality measurement in many U.S. IRFs (Fiedler, Granger, and Russel, 1998). Since January 1, 2002, items from the FIM™ are recorded in the IRF

patient assessment instrument (PAI) and, in combination with information on impairment, age, and comorbidities, used to assign Medicare patients to case-mix groups that determine the amount of payment under the IRF PPS. A different instrument, also including the FIM™, is recorded in Canadian inpatient rehabilitation. The minimum data set (MDS) (Hawes et al., 1995) is used for care planning and quality of care in U.S. skilled nursing facilities and in Canadian chronic care. Either MDS or the Medicare PPS assessment form can be used for payment purposes under the skilled nursing facility PPS. The Standardized Outcome and Assessment Information Set for Home Health Care (OASIS) is used for home care in the U.S. (Shaughnessey, Crisler, and Schlenker, 1997).

In this article, we focus on the role of functional status in classifying patients and, thereby, determining payment amounts in Medicare's IRF PPS. The assumption behind the inclusion of functional status in determination of payment amounts is that patients with lower function require additional resources. They will likely require a longer period of rehabilitation and/or more intensive therapy before they can return to the community. They likely require more nursing care each day they are in the hospital. If we provided the same payment independent of function, hospitals would have an incentive to discriminate against admitting patients with lower function. If admitted, the hospital might not have the resources to provide these patients with all needed treatment.

Grace M. Carter, Daniel A. Relles, and Gregory K. Ridgeway are with RAND. Carolyn M. Rimes is with the Centers for Medicare & Medicaid Services (CMS). The research in this article was supported by CMS under Contract Number 500-95-0056. The views expressed in this article are those of the authors and do not necessarily reflect the views of RAND or CMS.

We present analyses that show the relationship of the use of inpatient rehabilitation resources to level of functioning. In particular, we will show how the distribution of functional status varies with impairment. We will show that not all FIM™ items have the expected correlation with costs. Further, we will demonstrate the relationship between scales constructed from FIM™ items and cost and show that such scales can be used to construct groups that are homogeneous in resource use and suitable for case-mix adjustment for payment. The out-of-sample predictive validity and stability of these groups is covered elsewhere (Relles, Ridgeway, and Carter, forthcoming). Relles and colleagues used 4 years of data to show that groups constructed on each year's data predict quite well on the other 3 years of data, and explain approximately 90 percent of all variation in costs that can be explained using the FIM™ items.

We examine how incentives, administrative simplicity, and the potential for gaming affected the creation of case-mix groups for the IRF PPS. Administrative costs are another important consideration in using functional status for payment purposes. One of the driving factors in the development of the IRF PAI was to place only a reasonable administrative burden on hospitals for the collection and processing of data. The original versions of MDS and OASIS were criticized because of the required administrative burden. Less burdensome versions of each of these instruments are being implemented.¹

BACKGROUND

Rehabilitation hospitals and exempt units were excluded from the inpatient hospital PPS, which is based on diagnosis-related

¹The Medicare PPS assessment form, a shorter form of the MDS, was allowed as an option for assessments related only to payment beginning July 1, 2002. The new version of OASIS, labeled OASIS-B1(12/2002) was scheduled for implementation December 2002 according to CMS (Internet address: <http://cms.hhs.gov/oasis/default.asp>).

groups. The Tax Equity and Fiscal Responsibility Act continued to be the payment system for inpatient rehabilitation facilities because diagnoses alone inadequately captured resource use for these patients (Hosek et al., 1986). Thus, until the implementation of PPS for IRFs (January 1, 2002), the Medicare payment for rehabilitation was based on the actual cost compared with the target amount per case. This target amount was calculated from the historical costs trended forward.² A facility with operating costs below its target received its costs plus an incentive payment equal to the lower of 50 percent of the difference between the target and its costs or 5 percent of the target. New providers received Medicare costs for the first 3 years of operation (*Code of Federal Regulations*, 1996). The Tax Equity and Fiscal Responsibility Act contained no adjustments for the hospital's actual rehabilitation case mix or for the intensity of services required for different patient needs.

Measurement of Function in Rehabilitation Patients

In contrast to the acute inpatient hospitals payment system with its emphasis on medical conditions and treatments, and in contrast to long-term care with its emphasis on supportive and ameliorative care, rehabilitation care was and remains targeted to restoration of function:

“Disability occurs when functional limitations interfere with the performance of normal activities. Rehabilitation restores lost function and this restoration involves relearning motor and cognitive skills and transferring residual abilities into adaptive strategies.” (Stineman et al., 1994.)

Because rehabilitation restores function, any case-mix measurement system must measure the extent of the functional deficit

²The Balanced Budget Act of 1997 placed ceilings on the target amount.

to be restored. Thus, there was consensus regarding the importance of functional status for developing potential case-mix indicators and the need to go beyond existing administrative data systems (i.e., internal CMS data) but there was no agreement on the measures to be used (Stineman, 1995) or on data sources (Buchanan et al., 2002). A series of studies began to develop meaningful and reliable measures of functional status, and method and means for data collection.

Beginning in 1983, research focused on the development of a method for collecting functional status information. The initial step was the creation of a task force sponsored by the American Congress of Rehabilitation Medicine and the American Academy of Physical Medicine and Rehabilitation to develop a uniform data system for medical rehabilitation. This task force was created to meet the need to uniformly document the severity of patient disability and the outcomes of medical rehabilitation (George Washington University National Health Policy Forum, 1991). A grant was obtained from the Department of Education, and the National Institute on Disability and Rehabilitation Research to develop a MDS.

The task force reviewed existing scales (e.g., those measuring activities of daily living [ADLs]) and existing functional assessment instruments to select the most common and useful items for a rating scale to permit rehabilitation clinicians to assess severity of disability in a uniform and reliable manner. The Barthel Index (Mahoney and Barthel, 1965) and the Granger et al. (1986) modified version of this index served as the basis for what became the FIM™. The cumulative modifications to the Barthel Index include: addition of five items on communication and social cognition, increasing the three-level rating scale

to seven levels, and removing the weighting (Schoenman et al., 1991). From the deliberations of the task force a FIM™ instrument emerged.

The FIM™ measures functional status using 18 items covering six domains: self-care or ADLs (6 items on dressing upper and lower body, eating, grooming, toileting, and bathing), sphincter control (2 items on bowel and bladder management), mobility (3 transfer items), locomotion (2 items on walking/wheelchair use and stairs), communication (2 items on comprehension and expression), and social cognition (3 items on social interaction, problem solving, and memory). All 18 items are scored into one of seven levels of function ranging from complete dependence (level 1) to complete independence (level 7). The FIM™ motor and cognitive scores are Likert-like summated rating scales constructed from 13 and 5 FIM™ items, respectively. The motor scale covers the self-care, sphincter control, mobility, and locomotion domains and the cognitive scale covers the communication and social cognition domains.

The Guide for Uniform Data Set for Medical Rehabilitation (UDSmr³) was established for the assessment of functional status during medical rehabilitation. It includes demographic descriptions of the patient (birth date, sex, ZIP Code, ethnicity, marital status, living setting), clinical descriptions of the patient (impairment that is the primary reason for rehabilitation, *International Classification of Diseases, Ninth Revision, Clinical Modifications* (Centers for Disease Control and Prevention, 2003) diagnoses, functional independence measure at admission and discharge), and descriptions of the hospitalization

³ The UDSmr is operated by the Center for Functional Assessment Research of the State University of New York at Buffalo.

(encrypted hospital identifier, admission date, discharge date, payment source, and discharge living setting).

The next phase tested the instrument and analyzed data from it to develop case-mix measures. A number of studies tested the FIM™ instrument using the UDSmr database. For example, item difficulties were reported within impairment groups (Heineman et al., 1993). Linacre et al. (1994) showed that the 18 FIM™ items define two statistically and clinically different dimensions of motor and cognitive function. Stineman et al. (1996) used multi-trait scaling analysis to show that the simple summated motor and cognitive scales are internally consistent and have good convergent validity and discriminant validity.

Function-Related Groups

On a parallel research path, studies were assessing case-mix measurement groups to characterize the severity of a person's disability. In a CMS-funded study (Hosek et al., 1986) RAND and the Medical College of Wisconsin, using retrospective chart reviews, documented the statistical associations among functional status, rehabilitation, length of stay, and charges. Harada and colleagues (1993) used data collected from this study to develop functional related groups (FRGs). In this system the definition of case-mix groups started with assignment to one of nine rehabilitation-related conditions. These described the primary diagnosis for patients receiving rehabilitation, and functional status at the time of admission and/or changes in functional status. Because of the limited consistency in the methods that facilities used to enter these data, this study developed the feasibility of creating case-mix groups and the initial case-mix categories (Harada, 1991).

In a subsequent study, Stineman and colleagues (1994) developed the FIM™ FRGs using data from the FIM™ instrument. Patients were placed in groups beginning with an assignment to 1 of 17 rehabilitation groups (or an additional miscellaneous category) or rehabilitation impairment category (RIC) that describe the primary reason that the patient is receiving rehabilitation care. A subsequent analysis increased the number of RICs to 20 (Stineman et al., 1997). After assignment to an impairment category, patients were classified by functional status at admission. Functional status was described by the patient performance on the FIM™. These two concepts, impairment and functional status, were derived from two of the domains defined by the *International Classification of Impairments, Disabilities and Handicaps* (World Health Organization, 2001). Stineman (1997) noted that impairment is the anatomical defect, disease, or psychological state for which the individual is receiving rehabilitation and functional status measures the severity of disability. In the new ICF, impairment corresponds to the body function and structure component, and functional items measure the activity component within the environment of the rehabilitation facility.

A comparison of the FRGs with the FIM™-FRGs shows that the FIM™-FRGs were better predictors of length of stay (LOS). The FRG system explained 18.3 percent of the variance in LOS (Harada, Kominski, and Sofaer, 1993), but the FIM™ FRG explained 31 percent of the variation (Stineman et al., 1994). Subsequent work showed the FIM™-FRGs were similarly good predictors of other measures of resource use (charges and accounting costs). The functional measures in the FIM™ appeared to strengthen the predictive ability of FRGs to meet the assessment criteria:

“The degree to which the defined patient groups explain variation in resource use, the predictive gradient across groups within the system, the homogeneity of individual groups and the stability of predication in new data.” (Stineman, 1995.)

Case-mix measurement is a scientific and clinical process that identifies patient characteristics that predict outcomes of interest such as LOS, or the costs of an episode. The next step in the research was to assess the feasibility of using the FIM™ for payment. Because stakeholders have varying values, a payment system must include political and administrative considerations in addition to scientific ones. In 1994, CMS embarked on a series of studies to assess the feasibility of developing a PPS for IRFs using an updated version of the FIM™-FRGs to account for impairment and functional status (Carter, Relles, and Buchanan, 1997; Carter et al., 1997). These studies concluded that development of a PPS was feasible.

IRF PPS MANDATE

The law that governs the IRF PPS, as amended by the Balanced Budget Refinement Act of 1999, mandated the creation of classes of patient discharges or FRGs (referred to as a case-mix group) “based on impairment, age, comorbidities, and functional capability of the patient and such other factors as the Secretary deems appropriate to improve the explanatory power of functional independence measure-function related groups.” Each case-mix group was to be assigned “...a weighting factor that reflects the relative facility resources used for patients classified within the group as compared with patients classified within other groups...” (*Federal Register*, 2001). Payment rates are proportional to these weights.

DATA AND VARIABLE DEFINITIONS

Data Sources

Our primary data sources are the Medicare Provider Analysis and Review (MEDPAR) File which contains one record for each inpatient discharge paid by Medicare, the FIM™ data recorded by a subset of IRFs, and the annual cost reports from the Hospital Cost Report Information System. The FIM™ data come from the UDSmr and HealthSouth.⁴ In developing parameters for the IRF PPS, we used data from 1996-1999. In this article, however, we restrict our analyses to calendar year 1999 data.

The MEDPAR and FIM™ Files that described the same discharge were linked using a probability matching algorithm (Carter et al., 2002). The algorithm had two steps. The first step determined the Medicare provider number(s) corresponding to each facility code in the FIM™ database. The second step matched FIM™ and MEDPAR patients within paired facilities using a probabilistic match algorithm. In addition to hospital identity, the variables used were admission and discharge dates, ZIP Code, age at admission, sex, and race. All these variables are on each of the files, although sometimes in a slightly recoded form.

Estimating Cost

The Hospital Cost Report Information System Files contain information on costs and charges by cost center, facility characteristics, and utilization. Each record covers a hospital fiscal year. In the analyses reported here, we used the latest cost report for each hospital that was available

⁴ HealthSouth Corporation is a national provider of health services, with offices in Birmingham, Alabama.

in July 2000.⁵ We could not calculate costs for 2.4 percent of hospitals that were all-inclusive providers or otherwise were missing cost report data.

We used the departmental method to estimate the accounting cost of MEDPAR discharges. This method combines MEDPAR information about charges in each ancillary department with the departmental cost-to-charge ratio calculated from the cost report to estimate costs incurred by the patient in the department (Newhouse et al., 1989). Separate per diems for routine and special care days are combined with MEDPAR counts of such days to estimate routine and nursing costs. The per diems were inflated (or deflated) from the midpoint of the fiscal year to the day of discharge based on the observed rate of increase in hospital per diems (1.1 percent annually).

We use wage-adjusted cost per case as the dependent variable in our analyses. The wage adjustment affects 70.5 percent of costs, which is the labor share in the time period of our data. The hospital wage index used was prior to reclassification and reflects the elimination of teaching salaries.

Independent Variables

Our prediction of costs for cases discharged to the community is based on three sets of information: (1) RIC, (2) the 18 FIM™ items, and (3) patient age.⁶ The RIC is a grouping of codes that describe the impairment that is the primary cause of the rehabilitation hospitalization (Carter et al., 2002; *Federal Register*, 2001). The codes for the primary impairment are identical in the UDSmr and HealthSouth data. RICs

⁵ For 88.53 percent of discharges this was the cost report that began in Federal fiscal year (FY) 1998, for 11.22 percent of discharges in FY 1997, and for 0.25 percent of discharges in FY 1996.

⁶ The payment system also includes payment based on comorbidity.

were created based on clinical criteria and, except for the miscellaneous group, do not group patients who are clinically different from one another in the same RIC. We began with the 20 RICs defined in version 2 of the FRGs (Stineman et al., 1997). We evaluated these RICs and updated them to include an additional RIC for burns and changed the assignment of the multiple fracture codes (Carter et al., 2000).

In addition to using the individual FIM™ items as variables, we use the FIM™ cognitive scale (the sum of the items on communication and social cognition) and a modification of the motor score that will be explained in the results section. Age is taken from the MEDPAR and is age in years on the day of admission.

Sample Definition and Size

Table 1 shows that there were 390,048 discharges from IRFs in 1999. Of these, we were able to match FIM™ records for 257,024, or 66 percent of the MEDPAR population. Most of the unmatched MEDPAR records were from hospitals that did not participate in either of our FIM™ data sources. We judged the quality of the match, compared with what was possible given our data, in two ways. First, we looked at MEDPAR records for providers that appeared in a FIM™ database throughout 1999 and calculated the fraction of the MEDPAR records that we were able to match to a FIM™ record. We were able to match 90.1 percent of such MEDPAR records in 1999. The second way we judged the quality of the match is the percent of FIM™ records for which Medicare is listed as the primary payer that we were able to match. In calendar year 1999 we matched 95.9 percent of such FIM™ records. Using both measures, the match rate was very similar for each FIM™

Table 1
Medicare Discharges from Inpatient Rehabilitation Facilities (IRF) and Sample Sizes that Meet Various Criteria: 1999

IRF Category of Discharge	IRF Discharges
Population	390,048
Matched FIM™ Record Available	257,024
Complete FIM™ Data	256,702
Complete Cost and FIM™ Data	249,941
Exclude Transfers, Atypical Short Stays, Deaths, Age < 16, and >105, LOS > 365 Days	187,257
Excluding Outliers	186,766

NOTES: FIM™ is functional independent measure. LOS is length of stay.

SOURCES: Uniform Data System for medical rehabilitation; Medicare Provider Analysis and Review File.

source. Hospitals and cases in the analysis sample are reasonably representative of the population (Carter et al., 2002).

There were a small number of cases—roughly 0.1 of 1.0 percent of the sample—where the FIM™ data were incomplete. An additional 2.6 percent of matched cases were lost because we could not estimate case cost.

In this article we predict cost only for typical cases discharged to the community—i.e., excluding in-hospital deaths, transfer cases, and atypically short-stay cases. We used the MEDPAR verified date of death to identify the 0.5 percent of cases that died in the hospital. We used the FIM™ discharge setting variable to identify the 21.4 percent of sample IRF cases that were transfers. We also excluded the 2 percent of cases discharged to the community with LOS less than or equal to 3 days. LOS was taken from the MEDPAR. In addition, we excluded a handful of pediatric cases and cases with extremely high age or long LOS. Finally, we excluded less than 0.3 percent of cases whose estimated cost was outside a 3-standard deviation interval of the mean for the RIC on a log scale. Our final sample for the analyses presented here covers 186,766 discharges.

METHODS

Separate models of cost were fit within each RIC. For the purposes of this article we will provide summary information across

all RICs for each analysis. For the sake of brevity, we provide the details of some models only for the two largest RICs: stroke, and lower extremity joint replacement. (Additional details about other RICs are available on request from the authors.)

We use ordinary least squares regression (OLS) to examine the relationship between cost and individual FIM™ item responses. In an OLS model, a fixed amount of change in an independent variable, anywhere along its scale and no matter what the value of other variables, results in the same change in the prediction of the dependent variable. This is the simplest possible model and has parameters that are easily interpretable. It directly tests whether increasing functional independence is correlated with lower cost after controlling for other FIM™ items. Thus, it is a straightforward test of whether each item contributes to the prediction of cost in the expected way. For the items where this is not true, we looked more carefully at the FIM™ instructions and the consequences of including the variable in the FIM™ scales.

In order to demonstrate the relationship between cost and the FIM™ scales while controlling for age, we report the results of two models in addition to OLS. These are the generalized additive model (GAM) and classification and regression trees (CART). We construct graphs that allow us to interpret the results of these models.

The GAM approximates the relationship as a sum of smooth (rather than linear) functions of the independent variables (Hastie and Tibshirani, 1990). This means that a change in motor score from 20 to 21 might decrease predicted cost by a different percentage than a change from 60 to 61. GAM does not model interactions.

CART is the technique that was used to produce the FIM™-FRGs that are the basis of the case-mix groups used for payment in the IRF PPS (*Federal Register*, 2001). It is a well-known technique for building patient classification models (Breiman et al., 1984) and was used in the construction of the original FRGs. CART requires a dependent variable (log cost), and it seeks to develop predictors of the dependent variable through a series of binary splits from a candidate set of independent variables (age, FIM™ motor score, and FIM™ cognitive score). CART is invariant to one-to-one monotonic transformations of the independent variable such as those produced by Rasch (1980) analysis. CART partitions the data into two groups using the independent variables. Such a partition might separate patients with motor score exceeding 50 from those with motor score less than 50. CART chooses the variable on which to split the data and the value of the variable at which to split so that the new partitions minimize the squared prediction error. CART then recursively splits each partition until it satisfies some stopping criteria. As a result CART is invariant to one-to-one monotonic transformations of the independent variable so that an analysis using age or log (age) as an independent variable would produce the same model. This is particularly useful for handling the ordinal FIM™ motor and cognitive scores. CART's final product is a set of groups, each of which contains all patients with a specified range of the independent vari-

ables. Payments are then set to be proportional to the expected cost of all patients in the group.

RESULTS

Distribution of FIM™ Item Responses

Table 2 shows the mean and standard deviation of each of the 18 FIM™ items in our entire sample. For most items, the standard deviation is approximately 1.5—one-quarter of the six-point range of the item.

In order to use any functional measures in a payment system, we need to consider how formal and informal rules might affect patient classification. For example, items and scales for which many persons are placed at the bottom of the scale may be problematic because of the so-called floor effect—i.e., there may be real variation in the concept that the item or scale is attempting to measure that is not being captured. Similarly, a ceiling effect may conceal real variation at the top of the scale. In the FIM™ motor items (i.e., all but the last five items in the table), ceiling effects are apparently not a problem—eating is the item with the highest percentage of cases receiving the score of 7, and it is plausible that 40.9 percent of rehabilitation patients are, in fact, completely independent in eating. Similarly, it is plausible that 85.1 percent of rehabilitation patients are completely dependent at admission in going up stairs.

The remaining motor item with unusual data is transfer to tub or shower, where one-half of the patients were listed as completely dependent. On the surface, this appears strange, given that only 6.9 percent were completely dependent in transferring among bed and chair (or wheelchair) and only 12.2 percent in transfer to

Table 2
Mean and Standard Deviation of the 18 Functional Independent Measure (FIM™) Items and
Percent of Responses at Bottom and Top of Item Scale: 1999

Item	Mean	Standard Deviation	Level 1	Level 7
Eating	5.65	1.46	3.3	40.9
Grooming	4.92	1.25	2.2	13.3
Bathing	3.32	1.26	10.9	0.6
Dressing Upper Body	4.41	1.34	4.2	6.9
Dressing Lower Body	3.15	1.30	11.5	0.8
Toileting	3.55	1.45	12.0	2.3
Bladder Management	4.43	2.11	18.0	20.4
Bowel Management	4.79	1.73	8.1	12.0
Transfer to Bed or Chair	3.65	1.16	6.9	1.0
Transfer to Toilet	3.52	1.28	12.2	0.5
Transfer to Tub or Shower	2.35	1.51	50.1	0.2
Locomotion	2.31	1.46	39.9	0.3
Stairs	1.31	0.89	85.1	0.2
Comprehension	5.94	1.40	1.4	48.3
Expression	6.06	1.49	2.1	59.7
Social Interaction	5.99	1.44	1.6	53.9
Problem Solving	5.38	1.75	3.7	39.5
Memory	5.44	1.77	3.6	42.0

NOTES: Based on 186,766 discharges with complete values of admission FIM™. The FIM™ measures status using 18 items covering self-care or activities of daily living. All items are scored into one of seven levels of functioning ranging from complete dependence (level 1) to complete independence (level 7).

SOURCE: (Carter, G.M., Relles, D.A., Ridgeway, G.K., and Rimes, C.M., 2003.)

toilet. Based on informal conversations with hospital staff, we believe that a major reason for the high percentage of completely dependent scores on transfer to tub or shower is the UDSmr rule that, when an activity is not observed, then it is to be coded as 1 (completely dependent). This rule was formulated under the belief that the predominant reason why one of these 18 items was not observed would be because it was dangerous for the patient to try it. Such is completely plausible with the stairs item, for example. However, in discussing the transfer to tub rule with hospital staff, we found that many hospitals provide the patient with only sponge baths during the first 3 days of the stay; showers and tub baths are postponed until later in the stay. For such hospitals, the score of 1 in transfer to tub says nothing about the capability of the patient to perform this activity, and therefore nothing about the length or intensity of rehabilitation required.

The cognitive items individually show a potential for a ceiling effect. Given the complexity of cognitive functions such as

comprehension and expression, we cannot rule out the existence of a real ceiling effect from this data. Of course, even if there are cognitive levels of expression and memory that are not captured by the items, these may have little to do with resources required for rehabilitation.

Table 3 shows the mean value of the motor score minus the transfer to tub item and of the cognitive score for each RIC. Theoretically, the motor score varies from a 12 to 84. There is substantial variation across RICs, with the average motor score varying from 41.4 for traumatic brain injury to 49.8 for pulmonary patients. Low average values of the cognitive score are found only in stroke and brain injury RICs. There are a substantial number of cases at the cognitive ceiling in all other RICs.

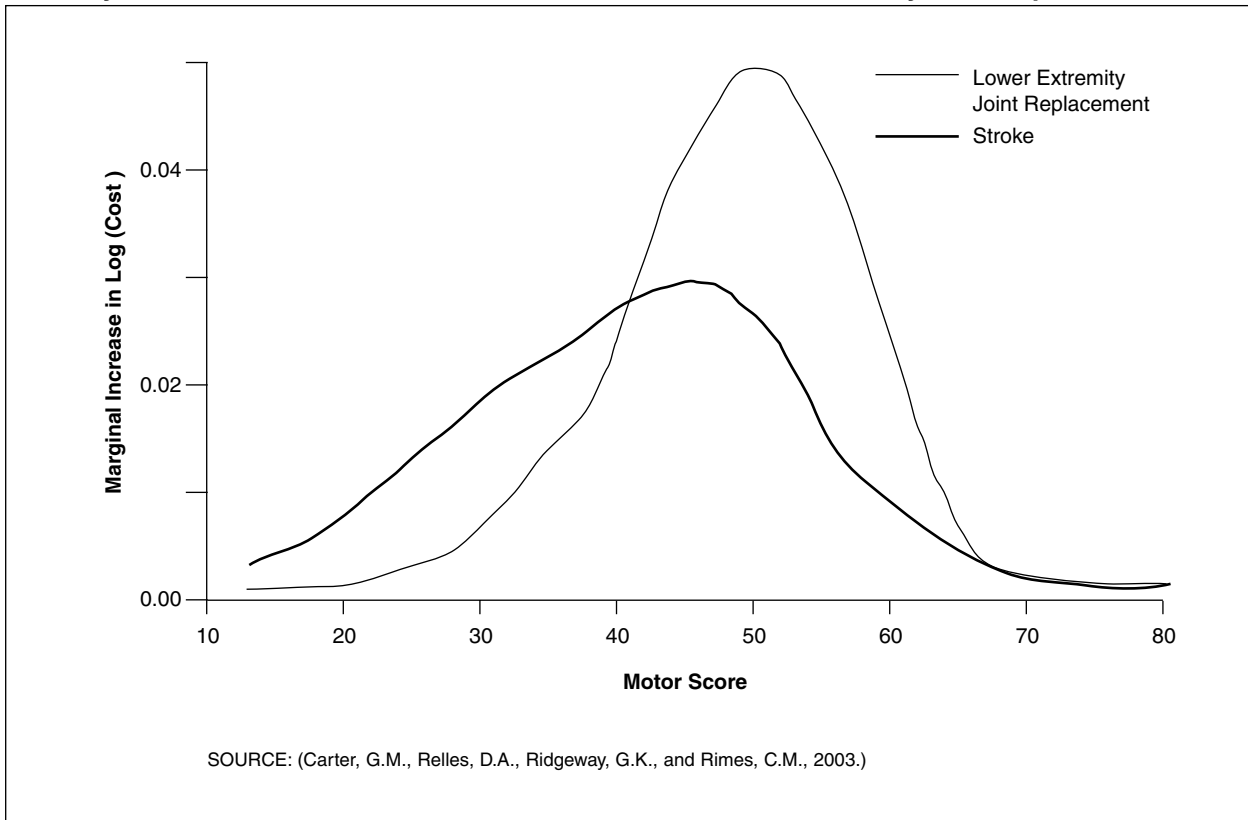
Figure 1 shows the distribution of the motor score minus the transfer to tub item for stroke and lower extremity joint replacement. There is almost a normal distribution in each RIC, but the stroke cases have a much larger standard deviation. Unlike the motor scores, the shape of the cognitive score distribution depends

Table 3
Mean Values of Motor and Cognitive Scores and Percent of Cases at Cognitive Ceiling, by
Rehabilitation Impairment Category: 1999

Category	Sample Size	Motor Score	Cognitive Score	Percent at Ceiling
Stroke	37,340	40.7	23.2	7
Traumatic Brain Injury	2,053	41.4	20.6	4
Non-Traumatic Brain Injury	3,758	41.9	21.7	6
Traumatic Spinal Cord	953	37.0	30.4	33
Non-Traumatic Spinal Cord	5,837	43.5	31.1	36
Neurological	8,875	43.0	27.6	18
Hip Fracture	20,627	42.9	29.8	29
Replacement of Lower Extremity Joint	43,427	48.6	32.5	49
Other Orthopedic	9,310	45.4	30.8	34
Amputation, Lower Extremity	6,156	45.7	30.6	33
Amputation, Other	662	47.3	30.6	33
Osteoarthritis	5,036	47.4	30.3	31
Rheumatoid, Other Arthritis	2,350	46.0	30.8	33
Cardiac	8,104	49.0	29.8	25
Pulmonary	5,382	49.8	30.1	28
Pain Syndrome	2,993	48.6	30.8	30
Major Multiple Trauma, No Brain or Spinal Cord Injury	1,679	41.2	29.9	30
Major Multiple Trauma, with Brain or Spinal Cord Injury	256	37.7	25.2	14
Guillain-Barre	313	40.7	31.1	31
Miscellaneous	21,553	45.2	28.7	22
Burns	102	42.7	27.7	19

SOURCE: (Carter, G.M., Relles, D.A., Ridgeway, G.K., and Rimes, C.M., 2003.)

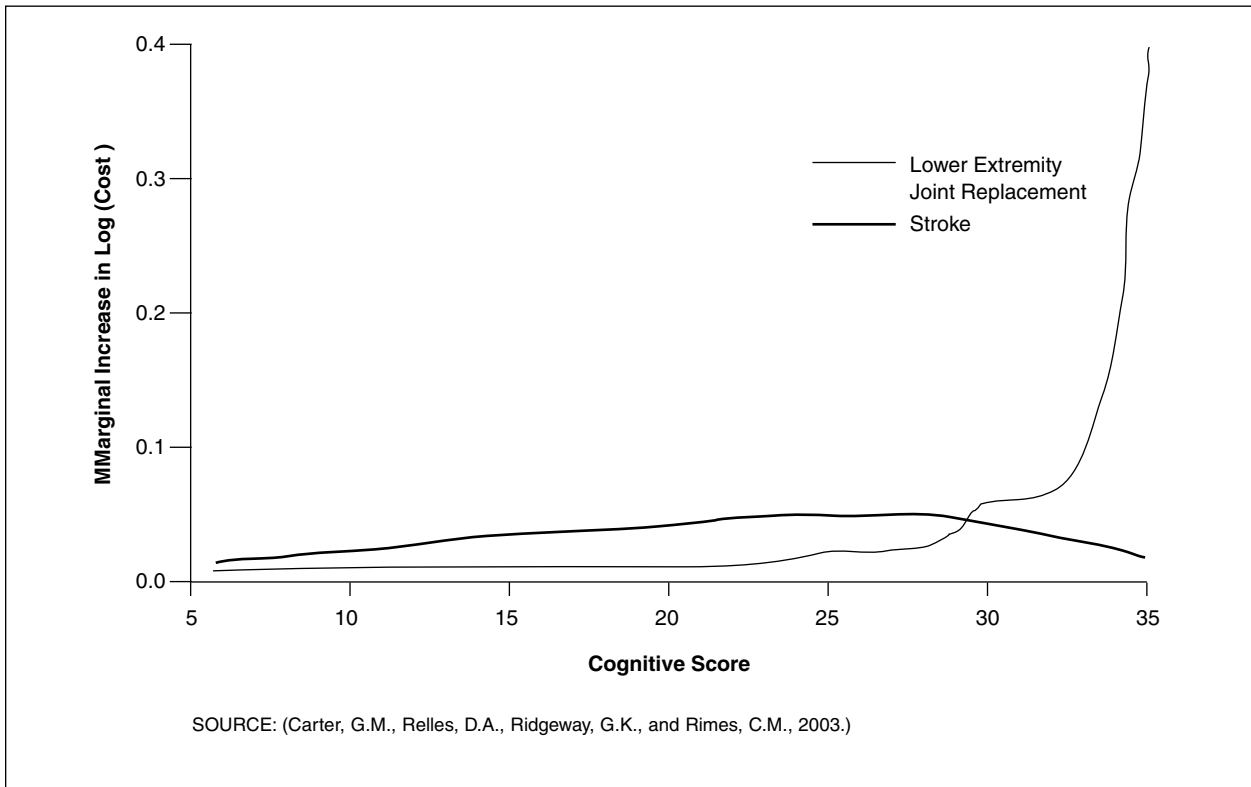
Figure 1
Density of Cases with Each Motor Score for Stroke and Lower Extremity Joint Replacement: 1999



SOURCE: (Carter, G.M., Relles, D.A., Ridgeway, G.K., and Rimes, C.M., 2003.)

Figure 2

Density of Cases with Each Cognitive Score for Stroke and Lower Extremity Joint Replacement: 1999



strongly on RIC (Figure 2). There is much more variation across cases in the stroke RIC.

Relationship Between Individual FIM™ Items and Cost

Table 4 shows the regression log of cost on each FIM™ item and age for each of six large RICs. For the range of values found in the table, the coefficient gives a good estimate of the percent increase in cost with an increase in one level of independence in the FIM™ scale. For example, the -0.030 coefficient on the eating item within the stroke RIC says that, all other responses and age equal, an increase of 1 in the eating item score results in a 3-percent drop in the expected cost of the case. The *t*-statistic shows the accuracy of the measurement of the coefficient with a *t*-statistic

with an absolute value of 2.0 or greater providing confidence that the coefficient is not 0 or of opposite sign from that shown here (statistically significantly different from 0 at $p < 0.05$).

Table 5 counts the coefficients from the regressions on all 21 RICs by their sign and range of value of *t*. Although the individual item effects are measured less precisely in the smaller RICs, the same items tend to have high likelihoods of the expected negative relationship between cost and independence and statistically significant *t*-statistics in both Tables 4 and 5.

Table 4 shows that in all six RICs there are substantial and significant decreases in cost with increasing independence in 7 of the 13 FIM™ motor items (eating, dressing lower body, toileting, bladder management, transfer to bed or chair, transfer to toilet, locomotion). The same is true in five

Table 4

Regression of Log of Cost on Functional Independent Measurement (FIM™) Items and Age Within Six Large Rehabilitation Impairment Categories (RICs): 1999

Item	Stroke		Neurological		Lower Extremity Fracture		Lower Extremity Joint Replacement		Cardiac		Miscellaneous	
	Coefficient	t-Statistic	Coefficient	t-Statistic	Coefficient	t-Statistic	Coefficient	t-Statistic	Coefficient	t-Statistic	Coefficient	t-Statistic
Eating	-0.030	-13.6	-0.032	-6.5	-0.015	-4.7	-0.007	-2.9	-0.043	-9.0	-0.037	-12.6
Grooming	0.021	6.5	0.008	1.2	-0.005	-1.5	-0.021	-8.5	-0.008	-1.3	0.007	1.8
Bathing	-0.031	-10.7	-0.007	-1.2	-0.027	-7.7	-0.027	-12.6	-0.033	-6.4	-0.017	-4.7
Dressing Upper Body	-0.019	-5.4	0.015	2.3	0.015	4.5	0.032	13.3	0.011	1.6	0.015	3.8
Dressing Lower Body	-0.050	-13.4	-0.044	-6.3	-0.046	-12.0	-0.046	-19.9	-0.051	-8.2	-0.044	-10.4
Toileting	-0.031	-10.5	-0.026	-4.4	-0.026	-8.1	-0.027	-13.5	-0.014	-2.8	-0.015	-4.4
Bladder Management	-0.012	-6.4	-0.011	-3.0	-0.014	-7.2	-0.010	-7.1	-0.015	-3.7	-0.013	-5.8
Bowel Management	0.010	4.9	-0.001	-0.1	0.001	0.5	0.000	-0.1	0.004	0.8	0.001	0.2
Transfer to Bed or Chair	-0.078	-21.0	-0.046	-6.3	-0.067	-16.2	-0.056	-20.2	-0.042	-6.1	-0.054	-11.8
Transfer to Toilet	-0.029	-8.2	-0.040	-5.6	-0.038	-9.7	-0.033	-14.3	-0.020	-3.5	-0.042	-9.8
Transfer to Tub or Shower	0.013	6.0	0.020	4.6	0.019	7.1	0.009	5.9	0.018	5.1	0.013	5.1
Locomotion	-0.057	-23.3	-0.037	-8.9	-0.037	-12.9	-0.032	-19.1	-0.042	-9.9	-0.034	-12.0
Stairs	-0.012	-3.9	-0.004	-0.6	-0.025	-4.1	-0.016	-5.6	-0.025	-4.4	-0.019	-4.7
Comprehension	0.028	10.3	0.036	5.1	0.012	2.7	0.043	10.9	0.013	1.8	0.018	3.8
Expression	-0.036	-15.7	-0.021	-3.2	-0.006	-1.2	-0.022	-4.8	-0.010	-1.3	-0.019	-4.1
Social Interaction	0.010	4.4	0.011	2.0	-0.004	-1.1	-0.024	-7.1	-0.012	-2.0	0.002	0.4
Problem Solving	-0.017	-5.5	-0.007	-1.0	0.003	0.6	-0.032	-8.2	-0.005	-0.7	0.011	2.2
Memory	-0.008	-2.7	-0.009	-1.4	-0.022	-4.7	-0.028	-7.1	-0.022	-3.1	-0.029	-6.2
Age	-0.003	-11.6	-0.001	-2.0	0.003	7.7	0.004	13.8	-0.003	-4.7	-0.005	-14.2
R ²	0.340	—	0.164	—	0.199	—	0.208	—	0.224	—	0.186	—
Number of Cases	37,340	—	8,875	—	20,627	—	43,427	—	8,104	—	21,553	—

NOTE: Negative coefficients mean that costs are higher for patients who are less independent.

SOURCE: (Carter, G.M., Relles, D.A., Ridgeway, G.K., and Rimes, C.M., 2003.)

Table 5

Counts of Coefficients, by Sign and *t*-Statistic Ranges in Regressions of Log Cost on Functional Independence Measure Items and Age Within Each Rehabilitation Impairment Category: 1999

Item	Negative Coefficients			Positive Coefficients		
	All	1	2	All	1	2
Eating	21	19	15	0	0	0
Grooming	9	5	1	12	7	2
Bathing	20	16	11	1	0	0
Dressing Upper Body	7	4	2	14	10	6
Dressing Lower Body	20	18	16	1	0	0
Toileting	20	19	13	1	0	0
Bladder Management	18	16	13	3	0	0
Bowel Management	9	5	3	12	5	2
Transfer to Bed or Chair	21	20	16	0	0	0
Transfer to Toilet	19	17	15	2	0	0
Transfer to Tub or Shower	3	1	0	18	16	14
Locomotion	21	19	18	0	0	0
Stairs	17	14	11	4	2	2
Comprehension	2	0	0	19	15	8
Expression	14	11	7	7	1	0
Social Interaction	12	7	3	9	3	2
Problem Solving	14	11	4	7	2	2
Memory	19	16	10	2	1	1

NOTE: Negative coefficients mean that costs are higher for patients who are less independent.

SOURCE: (Carter, G.M., Relles, D.A., Ridgeway, G.K., and Rimes, C.M., 2003.)

of the six RICs for two additional items (bathing and stairs). Table 5 shows that each of these 9 items was negative in between 17 and 21 of the regressions, and only stairs exhibited any positive and statistically significant coefficients.

One motor item, transfer to tub or shower, has consistently positive effects—costs increase with increasing independence. This is probably due in part to the over-coding of complete dependence, as previously discussed. However, it also may be due to the mixture of tub and shower in the same item and to the use of different types of assistive devices. For example, it may be that a patient who can transfer to a tub bench would need assistance in transferring to a shower seat. Thus, the transfer to tub/shower item provides only a situational measure of the person’s capabilities rather than an absolute measure. If patients with more capability were given harder situations at admission, this would help explain the positive and significant relationship between cost and transfer to tub or shower.

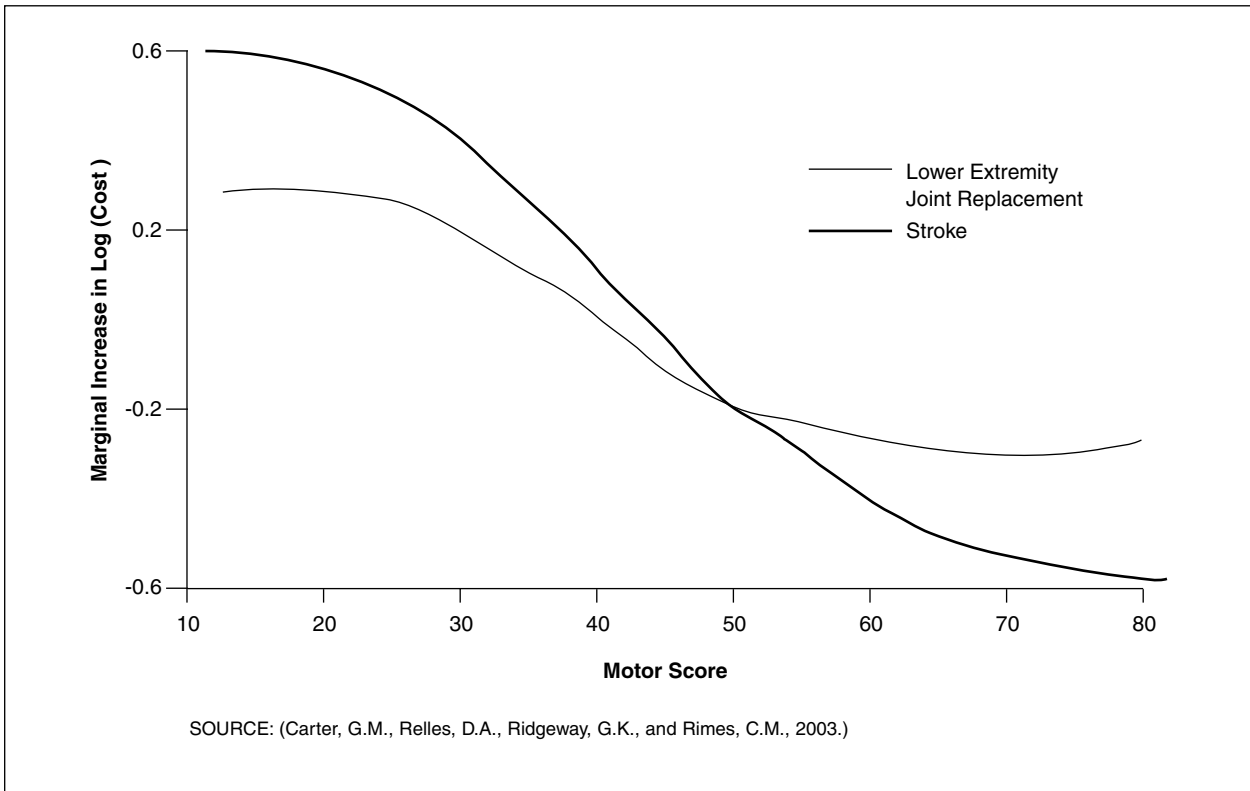
The remaining three motor items—bowel management, dressing upper body, and grooming—appear to predict cost in some RICs, but not in others. Although we do not understand this completely, it may be relevant that bowel management, transfer to tub/shower, and dressing upper body had the lowest reliabilities of any FIM™ items in a recent study (Buchanan et al., 2002).

Although the cognitive items do not predict cost as consistently as the motor items, two of the items (expression and memory) are negative and statistically significant predictors of cost in either all or five of the six large RICs (Table 4). In addition, social interaction and problem solving, which each significantly predict cost in some of the RICs (Table 4), are much more likely to have negative coefficients with substantial values of *t* than to have positive coefficients with substantial values of *t* (Table 5).

However, increasing independence in the comprehension item is positively related to cost in each of the large RICs and in

Figure 3

Motor Score Component of the Generalized Additive Model Fit for Stroke and Lower Extremity Joint Replacement: 1999



many others as well. It may be that many hospitals do more for patients that understand what is happening, or that such patients can tolerate more therapy.

To confirm that OLS was not overlooking important non-linear effects, we looked at plots of the marginal contributions of each component as estimated by GAM. Although log cost did not always smoothly decline with the seven-point scale, there were only two items where the relationship was consistently perverse. These were the same ones that showed up in the linear models: comprehension and transfer to tub. Likely due to the situational nature of the transfer to tub item, in many large RICs, cost increased with increasing independence in the higher values of independence as well as the lowest.

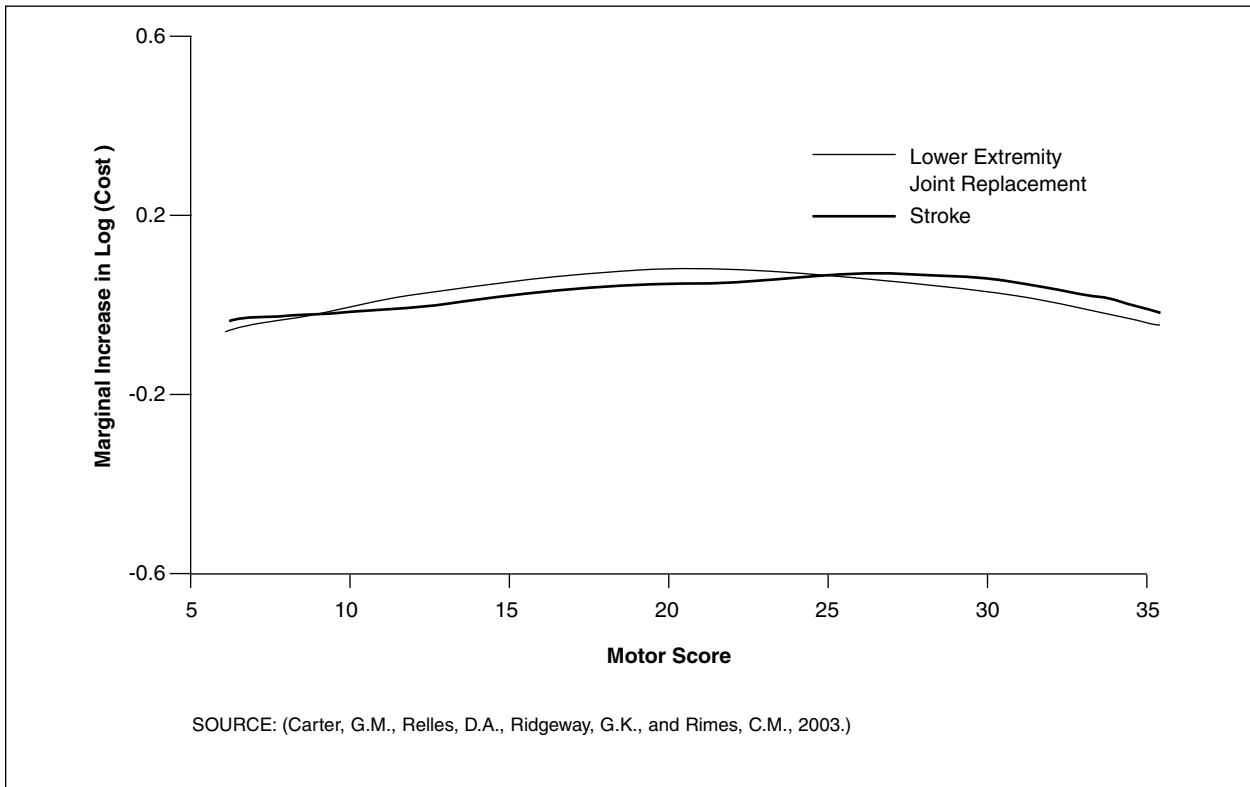
Relationship Between FIM™ Scales and Cost

We use GAM to show the marginal contribution of motor and cognitive scales to the estimated log cost. OLS coefficients provide marginal estimates, but they enforce linear effects. GAM provides marginal estimates and allows arbitrary curvature. We define the scales as they are used in the IRF PPS, dropping transfer to tub from the motor score.

We found that cost declines smoothly with increases in function as measured by the modified motor score in each of the large RICs (and throughout most of the range in all RICs). Figure 3 uses the two largest RICs to illustrate our results. The motor effects are large and sloping in the

Figure 4

Cognitive Score Component of the Generalized Additive Model Fit for Stroke and Lower Extremity Joint Replacement: 1999



expected direction (larger scores yield lower costs). The total decline in cost throughout the range of motor scores is slightly higher in stroke than in joint replacement. In each RIC, there is an area of low motor function where the decline in costs is modest, but the decrease in cost with each unit increase in independence accelerates and becomes substantial. For example, at the median motor score in stroke (42), a 1-point increase in independence is associated with a 3.3-percent decrease in costs (after controlling for age and cognitive score). There is also a region at the upper motor score where the rate of decline in costs with increasing motor score slows substantially.

The cognitive effects are shown in Figure 4. These tend to be much smaller, very close to zero. Unlike the motor score, costs do not decline uniformly with the

cognitive score. For lower values of the cognitive scale, higher scores are associated with higher costs.

Case-Mix Groups

Within each RIC, CART was used to create groups that meet the IRF PPS mandate—specifically groups defined by age, modified motor score, and cognitive score that are relatively homogeneous with respect to resource use. The groups were subsequently divided based on comorbidity tiers.⁷

Certain considerations beyond the ability to predict cost entered into the decisions that created the case-mix groups. CMS decided that the groups should be defined so that they have monotone weights in the

⁷ Additional case-mix groups are defined for atypically short-stay cases and for in-hospital deaths. There are also special payment provisions for transfer cases and high-cost outlier cases.

FIM™ scores—i.e., that if two patients are in different groups and differ only on one scale, then the hospital should receive a higher payment for the patient with the lower function. This is consistent with the assumption that patients with lower function often require additional resources. By maintaining or increasing payment for patients with lower function, hospitals should have the resources to provide these patients with needed treatment and should have no reason to discriminate against such patients at admission.

The FIM™ scales used in the creation of the CMGs were the sum of the 12 FIM™ motor items excluding transfer to tub/shower and the sum of all 5 cognitive items. We recommended that these scales be chosen by CMS after analysis and on the advice of our technical expert panel.⁸ In addition to the analyses presented above, we compared case-mix groups created using the original and modified motor score and using the cognitive scale and one that dropped comprehension. Relles, Ridgeway, and Carter (forthcoming) show that the index without transfer to tub was a slightly better predictor of cost than the index with it in all combinations of fitting year and prediction year. The situational nature of the item might allow hospitals to game their response. The technical expert panel agreed that transfer to tub/shower should not affect payment in the form in which it appears on the FIM™.

We also analyzed dropping comprehension from the cognitive scale because its relation to cost is opposite to that of the standard cognitive scale in which it is embedded. After fixing a stopping rule, dropping comprehension from the index produces a slightly better prediction in some years. However, eliminating comprehension raises issues related to incen-

⁸ The panel consisted of 22 clinicians, researchers, and IRF administrators. Names and affiliations are found in Carter et al. (2002).

tives and fairness. Because the cognitive scale has only a weak relationship to cost, it is used only occasionally in the definition of CMGs. Dropping comprehension does not increase the frequency with which FRGs are defined by cognitive function. When the full cognitive scale is used, the other four items determine the direction of the cognitive effect so that a higher cognitive score results in a lower payment when it has any effect at all. We could eliminate splits that contradict this general result if they were to occur. If we take the comprehension item out of the index, the system will provide no extra incentives to treat patients with lowered comprehension. If some hospitals do spend extra to treat such patients, they would not be compensated for such extra resources. The improvements in predicting cost are so slight that it seemed to us that the decision should be based on clinical judgment about what should be paid for. Based on the advice of our technical expert panel, we recommended keeping the comprehension item in the cognitive score.

CART attempts to replicate the patterns shown in Figures 3 and 4. If there were discontinuities in these curves, CART could exploit them to explain the variation. However, the cost curves are continuous, and CART can only approximate a smooth curve by a series of discrete jumps. This presents a tradeoff related to the number of such jumps. For ease of administration, it is convenient that there not be too many groups. A very small number of groups allows substantial differences in payment at boundary points in the FIM™ scale and thus, larger incentives for coding creep. In our data set, using some published stopping rules, we could have created hundreds of groups. Instead, we used a stopping rule that limited the number of groups produced by CART to approximately 100. CART can find interactions

such as a greater importance of the cognitive scale at higher values of the motor scale.

Cost is strongly influenced in the expected direction by functional status as measured by the motor scores. Thus, CART largely splits on motor scores and almost all splits reflect increasing cost with decreasing motor score. On the other hand, the cognitive effects are relatively flat and occasionally not monotone. This is true for an index that drops comprehension as well as the standard cognitive index.

In order to create only a manageable number of splits, we used a stopping rule that placed confidence bands around the cross-validated estimate of prediction error by tree size (Breiman et al., 1984). This rule stops partitioning when the prediction error is within one standard error of the minimum. In our data this reduces the number of groups to approximately 100. It should also reduce the probability of overfitting, and could cause some more heterogeneous groups (in terms of log cost) to be combined.

We took the CART output and modified it to reflect the need for monotonicity by joining groups that were not monotonic. We also forced groups that differ on a single factor (i.e., adjacent nodes of a tree) to differ by more than \$1,500 in payment amount by joining adjacent bottom nodes of a tree.

Information showing the regression tree for stroke and other large RICs is available from the author upon request. These simple trees identify groups with widely different costs. For example, the stroke patients with the highest set of motor and cognitive scores typically cost only \$5,064, while the group with the lowest motor score cost and age 82 and under cost \$20,869, or four times as much as the least expensive group.

DISCUSSION

Functional status is an important predictor of the use of resources in inpatient rehabilitation. We saw strong relationships between cost and the FIM™ motor score whether we used ordinary regression, GAM, or CART. Most FIM™ motor items individually contribute to predicting cost in all large impairment groups, and all but one contribute in some RICs.

In order to use any functional measures in a payment system, we need to consider how formal and informal rules might affect patient classification. Buchanan et al. (2002) emphasizes the importance of details in the construction of functional items to be used for payment. The analyses again demonstrate the extent to which details matter. The transfer to tub item did not show the expected correlation with cost for two reasons. First, the situational nature of the item which makes it subject to gaming. Although conversations with clinicians and a perusal of the IRF PAI instructions did not reveal any other FIM™ items with similarly large situational dependencies, we cannot be sure that none exist. CMS clarified instructions for the FIM™ items on the IRF PAI, but it is likely that not all problems were addressed. The dressing items are particularly suspect because of uncontrolled variation in the type of garment being used.

The second reason for the anomalous correlation of transfer to tub with cost is the FIM™ rule which scores unobserved items as most dependent. Although this rule applies to all items, it is a problem only when there are reasons other than capability that affect the likelihood of being unobserved. The distribution of item scores presented here suggests that transfer to tub was the only item where unobserved occurred frequently enough to pose a problem for payment.

The new IRF PAI will let us find out how frequent the problem of unobserved is and how we should treat these cases in creating a functional independence scale. This information may improve the predictive ability of all the FIM™ items. The new form also provides for separate scoring of tub and shower transfers and, together with the unobserved flag, may allow transfer to tub/shower back into the payment system.

We also found that poor comprehension does not predict increased cost. For most RICs, the cognitive scale, in its entirety or dropping comprehension, does not predict increased costs. Further, we found that, for our large RICs at low levels of the total cognitive scale, resource use decreases with lowered cognitive function. At this point it is not clear if this is a measurement problem or if, instead, practice patterns do not provide more resources for those with greater cognitive deficits. Poor cognitive performance might limit ones ability to benefit from rehabilitation. Because the reason(s) for the empirical relationships of cost with comprehension and low values of the cognitive scale are not known, CMS opted to insist that payments for patients with lower comprehension or cognitive scale be no lower than for otherwise similar patients. This decision provides the greatest protection for such patients and for hospitals that care for them.

The items that best measure differences in functional deficit between members of a population will vary with the diseases found in the population and with demographics of a population. For example, the distribution of the FIM™ cognitive scale is quite different between stroke patients and orthopedic patients. Because functional deficits vary across disease groups, the best predictions of resource use probably use different ADLs for different groups,

and indeed we found differences in the OLS coefficients for the same item across RICs.

Nevertheless, two simple scales—the reduced motor scale and the cognitive scale—can be constructed from a very short, easily administered instrument. These simple scales predict resource use well within the varied inpatient rehabilitation population. FRGs constructed from these scales predict costs over time and out of sample. In Relles, Ridgeway, and Carter (forthcoming), we show that 100 FRGs explain roughly 81 to 85 percent of the variance explained by gold standard models that use more detailed scales, depending on year. If we compare the CART model with the gold standard models that use only the same indices as CART (instead of more information), we explain more than 90 percent of the explainable variance. In the same article, we compared actual and predicted FRG means for non-fitting years in each RIC and found them to be quite close. The actual means in adjacent FRGs were also well separated in the non-fitting years.

FUTURE WORK

Because of the variation in the importance of different functions in different RICs, it is possible that using different summation of FIM™ items in different RICs, instead of using the modified motor and cognitive scales in each RIC, would improve prediction of cost. Further, new information might help. For example, other dimensions of cognitive performance such as executive function or motivation or depression might improve the cognitive scale. For example, Eilertsen et al. (1998) found that the mini mental status exam together with FIM™ motor score and a measure of depression behaviors had

substantially greater explanatory power for stroke cases than age, motor, and cognitive scales as used in the FIM™-FRGs. Instrumental ADLs or measures of cognitive function might better predict resource use in the orthopedic groups or in all patients with high scores on the FIM™ scales. Similarly, additional questions about orientation, memory, and personal interactions might help predict resource use for those with low cognitive score.

Uncalibrated instruments with different scales and/or different definitions for functional measures make it difficult to transfer information as the patient moves across different settings—a primary process if we are to improve quality of care (Institute of Medicine, 2001). But we also need to use an instrument that is non-burdensome and contains information that is adequate for payment. The IRF PAI was created to maintain the administrative simplicity of the FIM™ while providing additional information to improve case-mix groups in the immediate future.

In the longer run, we need to augment or replace the IRF PAI to produce a measurement tool or tools that contain uniform information across sites (Medicare Payment Advisory Commission, 1999). Such an instrument must contain enough information to adequately predict resource needs in individual site types, including consistent measures of ADLs and cognitive function. Judicious use of skip patterns should help to keep the burden under control. For example, instrumental ADLs might be assessed only for those with high cognitive function and details of cognitive deficits only for those with apparent need. Such a screen is used to trigger collection of additional information about communication, financial management, and orientation in the Canadian rehabilitation MDS (Canadian Institute for Health Information, 1999).

REFERENCES

- Breiman, L. J., Friedman, M., Olshen, R.A., and Stone, C.J.: *Classification and Regression Trees*. Wadsworth, Inc. Belmont, CA. 1984.
- Buchanan, J. L., Andres, P., Haley, S., et al.: *Final Report on Assessment Instruments for PPS*. RAND. Santa Monica, CA. 2002.
- Carter, G. M., Relles, D.A., and Buchanan, J.L.: *A Classification System for Inpatient Rehabilitation Patients: A Review and Proposed Revisions to the Functional Independence Measure—Function Related Groups*. National Technical Information Service Number: PB98105992. RAND. Santa Monica, CA. 1997.
- Carter, G. M., Buchanan, J.L., Donyo, T., et al.: *A Prospective Payment System for Inpatient Rehabilitation*. RAND. National Technical Information Service Number: PB98106024. Santa Monica, CA. 1997.
- Carter, G. M., Buntin, M.B., Hayden, O., et al.: *Analyses for the Initial Implementation of the Inpatient Rehabilitation Facility Prospective Payment System*. RAND. Santa Monica, CA. 2002.
- Carter, G.M., Relles, D.A., Ridgeway, G.K., and Rimes, C.M.: *Measuring Function for Medicare Inpatient Rehabilitation Payment*. Interim Report to the Centers for Medicare & Medicaid Services. RAND. Santa Monica, CA. 2003.
- Carter, G. M., Relles, D.A., Wynn, B.O., et al.: *Interim Report on an Inpatient Rehabilitation Facility Prospective Payment System*. RAND. Santa Monica, CA. 2000.
- Centers for Disease Control and Prevention: *International Classification of Diseases, Ninth Revision, Clinical Modification (ICD-9-CM)*. Internet address: <http://www.cdc.gov/nchs/about/otheract/icd9/abtcd9.htm>. (Accessed 2003.)
- Code of Federal Regulations*: Title 42-Public Health, Vol. 2, Part 413.30(d). Office of the Federal Register, National Archives and Records Administration. U.S. Government Printing Office. Washington, DC. October 1, 1996.
- Eilertsen, T. B., Kramer, A.M., Schlenker, R.B., and Hrinkevich, C.A.: Application of FIM™-FRGs and RUGs-III Systems Across Post Acute Settings. *Medical Care* 36(5):695-705, May 1998.
- Federal Register*: Medicare Program; Prospective Payment System for Inpatient Rehabilitation Facilities; Final Rule. 66FR, 41315-41430, August 7, 2001.

- Fiedler, R. C., Granger, C.V., and Russel, C.F.: Uniform Data System for Medical Rehabilitation. *American Journal of Physical and Medical Rehabilitation* 77(5):444-450, September/October 1998.
- George Washington University National Health Policy Forum: *An Update on Functional Assessments: Perspectives from Consumers, Practitioners, Policy Makers and Payers*. Washington, DC. May 1, 1991.
- 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*. Aspen Publishing Company. Baltimore, MD. 1986
- Harada, N. D.: *The Development of a Resource-Based Patient Classification Scheme for Rehabilitation*. University of California Los Angeles. Los Angeles, CA. 1991.
- Harada, N. D., Kominski, G., and Sofaer, S.: Development of a Resource-based Patient Classification Scheme for Rehabilitation. *Inquiry* 30(1):54-63, Spring 1993.
- Hastie, T. J., and Tibshirani, R. J.: *Generalized Additive Models*. Chapman and Hall. New York, NY. 1990.
- Hawes, C., Morris, J.N., Phillips, C.D., et al.: Reliability Estimates for the Minimum Data Set for Nursing Home Resident Assessment Care Screening. *The Gerontologist* 35(2):172-178, April 1995.
- Heineman, A.W., Linacre, J. M., Wright, B.D., et al.: Relationship Between Impairment and Physical Disability as Measured by the Functional Independence Measure. *Archives of Physical Medical Rehabilitation* 74(6):566-573, June 1993.
- Hosek, S., Kane, R., Carney, M., et al.: *Charges and Outcomes for Rehabilitative Care: Implications for the Prospective Payment System*. RAND. Santa Monica, CA. November 1986.
- Institute of Medicine: *Crossing the Quality Chasm: A New Health System for the 21st Century*. National Academy Press. Washington, DC. 2001.
- Linacre, J. M., Heineman, A.W., Wright, B.D., et al.: The Structure and Stability of the Functional Independence Measure. *Archives of Physical Medical Rehabilitation* 75(2):127-132, February 1994.
- Mahoney, F.I., and Barthel, D.W.: Functional Evaluation: The Barthel Index. *Maryland State Medical Journal* 14:61-65, 1965.
- Medicare Payment Advisory Commission: *Report to the Congress: Medicare Payment Policy*. Washington, DC. March 1999.
- Newhouse, J. P., Cretin, S., and Witsberger, C.J.: Predicting Hospital Accounting Costs. *Health Care Financing Review* 11(1):25-33, Fall 1989.
- Rasch, G.: *Probabilistic Models for Some Intelligence and Attainment Tests*. University of Chicago Press. Chicago, IL. 1980.
- Relles, D. A, Ridgeway, G.K., and Carter, G.M.: Data Mining and the Implementation of a Prospective Payment System for Inpatient Rehabilitation. *Health Services and Outcomes Research*. Forthcoming.
- Schoenman, J., McLaughlin, B., Stone, R., and Griffiths, S.: *Identification and Evaluation of Patient Classification Systems for PPS Excluded and Non-PPS Providers: Final Report*. Prepared for the Payment Assessment Commission. Project Hope Center for Health Affairs. January 1991.
- Shaughnessey, P.W., Crisler, K., and Schlenker, R.: *Medicare's Oasis: Standardized Outcome and Assessment Information Set for Home Health Care; OASIS-B*. Center for Health Services and Policy Research. Denver, CO. 1997.
- Stineman, M. G., Escarce, J. J., Goin, J. E., et al.: A Case-Mix Classification System for Medical Rehabilitation. *Medical Care*. 32(4):366-379, April 1994.
- Stineman, M. G.: Case Mix Measurement in Medical Rehabilitation. *Archives of Physical Rehabilitation* 76(12):1163-1170, 1995.
- Stineman, M. G., Shea, J., Jette, A., et al.: The FIM™: Tests of Scaling Assumptions, Structure and Reliability Across 20 Diverse Impairment Categories. *Archives of Physical Rehabilitation* 77(11):1101-1108, November 1996.
- Stineman, M. G., Tassoni, C.J., Escarce, J. J., et al.: Development of Function-Related Groups, Version 2.0: A Classification System for Medical Rehabilitation. *Health Services Research* 32(4):529-548, October 1997.
- Stineman, M. G.: Measuring Case Mix, Severity, and Complexity in Geriatric Patients Undergoing Rehabilitation. *Medical Care* 35(6):JS90-JS105, Supplement 1997.
- World Health Organization: *International Classification of Functioning, Disability, and Health*. World Health Organization. Geneva, Switzerland. 2001.
- World Health Organization: *International Classification of Impairments, Disabilities and Handicaps*. World Health Organization. Geneva, Switzerland. 2001.

Reprint Requests: Grace M. Carter, RAND, 1700 Main Street, P.O. Box 2138, Santa Monica, CA. 90407-2138. E-mail: grace_carter@rand.org