TECHNICAL
REPORT

Possible Refinements to the Construction of Function-Related Groups for the Inpatient Rehabilitation Facility Prospective Payment System

Daniel A. Relles, Gregory K. Ridgeway,
Grace M. Carter, Melinda Beeuwkes Buntin

Supported by the Centers for Medicare and Medicaid Services

RAND HEALTH

The research described in this report was supported by the Centers for Medicare and Medicaid Services.

Published 2005 by the RAND Corporation
1776 Main Street, P.O. Box 2138, Santa Monica, CA 90407-2138
1200 South Hayes Street, Arlington, VA 22202-5050
201 North Craig Street, Suite 202, Pittsburgh, PA 15213-1516
RAND URL: http://www.rand.org/
To order RAND documents or to obtain additional information, contact
Distribution Services: Telephone: (310) 451-7002;
Fax: (310) 451-6915; Email: order@rand.org

Preface

This report covers the RAND Corporation's analyses concerning the possible refinement of the comorbidity parameters of the Inpatient Rehabilitation Facility Prospective Payment System (IRF PPS). This work was performed during phase II of our project to design, develop, implement, monitor, and refine the IRF PPS. Phase II began in October 2001. Implementation of the IRF PPS began January 1, 2002. This research was sponsored by the Centers for Medicare and Medicaid Services (CMS) under contract No. 500-95-0056 and is one part of the final report on that project. The research was conducted within RAND Health, a division of the RAND Corporation. A profile of RAND Health, abstracts of its publications, and ordering information can be found at www.rand.org/health.

For some analyses in this report, we estimated the costs of IRF services provided to Medicare beneficiaries in 2002. To make these estimates, we generally used cost report information in the public use files that matched the date of the beneficiary's discharge (in other words, for a beneficiary discharged June 1, 2002 we used the IRF's cost report that included June 1, 2002, assuming it was available on the file). After this report was completed, but during the public comment period on the proposed rule updating the IRF PPS effective October 1, 2005, HealthSouth, a large chain organization, notified CMS that its IRFs did not include any home office costs in their cost reports for cost reporting periods beginning on or after October 1, 2001 and before October 1, 2003. Home offices of chain organizations such as HealthSouth usually furnish central management and administrative services such as centralized accounting, purchasing, personnel services, management, and other services to support patient care services furnished by its member providers. The reasonable costs of these services are normally included in the provider's cost report and reimbursed as part of the provider's costs. The home office costs for HealthSouth are approximately 13 percent of total costs for its IRFs. The home office costs were included in the cost reports used to estimate 1999 costs for HealthSouth IRFs but were omitted from their cost reports covering 2002 discharges. The HealthSouth hospitals cared for about 19 percent of the cases in our sample hospitals and we estimate that analyses in this report are based on costs per case that were understated by approximately 1.6 percent on average, and by about 6 percent for freestanding IRFs. For further information on this issue, see the IRF PPS final rule (Department of Health and Human Services, Centers for Medicare and Medicaid Services, "Medicare Program; Inpatient Rehabilitation Facility Prospective Payment System for FY 2006; Final Rule," *Federal Register,* Vol. 70, No. 156, August 15, 2005, p. 47884).

Contents

APPENDIX

Figures

Tables

Executive Summary

Purpose and Approach

In 2002, Medicare implemented a prospective payment system (PPS) for inpatient rehabilitation facilities (IRFs). The PPS works by assigning patients to groups according to how well patients function. These groups, called function-related groups (FRGs), are then used to predict the cost of treating particular Medicare patients according to their ability to function in four general categories: transfers, sphincter control, self-care (e.g., grooming, eating), and locomotion. Patient functioning is measured according to 18 categories of activity—13 motor tasks, such as climbing stairs, and 5 cognitive tasks, such as recall.

As part of a contract to monitor how accurately the IRF PPS is predicting treatment costs, the Center for Medicare and Medicaid Services (CMS) asked RAND to examine possible refinements to the FRGs to identify potential improvements in the alignment between Medicare payments and actual hospital costs. Several developments make it likely that significant refinements can be made.

- When the IRF PPS was implemented in 2002, a new recording instrument was used to collect patient data. Known as the IRF Patient Assessment Instrument (or IRF PAI), the new instrument contained questions that improved the quality of the information available to us.
- We also have more recent data on a larger patient population: Instead of 1999 data on rehabilitation patients from just a sample of hospitals, we now have 2002 data that describe the entire universe of rehabilitation patients.
- We have implemented improvements in the algorithms that produced the initial FRGs, which should improve prediction of treatment costs.
- In addition, two years have passed since the initial FRGs were created, and changes in the cost structure of IRFs have occurred.

Our analysis had two specific objectives: (1) to explore whether the new data enable better prediction of treatment costs and (2) to assess possible refinements to the FRGs based on the new data.

To address the first objective, we performed two tasks to examine the new items from the data set, on IRF patient functioning and costs. Specifically, we reexamined assumptions about whether particular indicators that an activity was not observed, or "missing," indicated a lack of functioning or simply absent data. We also looked at the usefulness of some new indicators in the IRF PAI data for predicting costs.

To address the second objective, we also performed two tasks: First, we considered whether alternative indices that included weighting for patient functioning might predict costs more accurately; second, we ran the algorithm used in 1999 to derive FRGs with the new IRF PAI data to see whether the FRGs would look substantially different.

Key Findings

Missing indicators. The earlier data did not make a distinction between patient dependence on others in performing some function and a missing report about that function. In effect, the earlier data assumed that no report about a particular function meant that patients were unable to perform it. The new IRF PAI data allowed us to revisit this assumption. We found that, for most activities, this assumption held true. However, for two particular activities—"transfer to toilet" and "transfer to tub"—it did not. A lack of data on these activities should be interpreted less strongly than for the other missing indicators. We flag this issue for future consideration in payment adjustment.

Importance of "function modifiers." Another set of variables contained in the IRF PAI data are "function modifiers," which provide more-nuanced information about patient functioning than do the basic measures (the so-called functional independence measurements, or FIM™ items). For example, one such modifier is distance walked, which adds information to the basic FIM™ category "walking."

We found that this type of additional information was useful in predicting costs for two conditions: bowel or bowel incontinence (proxied by a question eliciting "frequency of accidents," not further defined). Incontinent patients cost more, although we could not determine the degree to which high-frequency or low-frequency incontinence made an additional difference.

We also found that one of the function modifiers—distance walked—might actually be a better FIM™ category than simply "walking." We think it is worth considering whether the FIM™ measurement could be improved so that its explanatory power is at least that of distance walked.

Indices and weighting. Currently, there are two summary measures that capture the two dimensions of the FIM™ measures: One summarizes motor functioning; the other summarizes cognitive function. In these indices, the different measures are weighted equally. We wanted to explore whether assigning weights to the measures would help improve prediction of treatment costs. Our results suggested that we might expect some improvement in explanatory power by using a motor index that does not equally weight all components.

Refinements to the FRGs. We reran the algorithm that produced the 1999 FRGs, but using the 2002 data. We found that the 2002 runs produced many fewer payment groups across the various conditions—83, versus 95 before—so the payment groups look somewhat different. The main differences were in stroke, which had 14 groups before but only 8 now. For most of the conditions, costs are driven simply by motor score. The exceptions are stroke and traumatic spinal cord injury, for which younger people cost more, perhaps because they can take more therapy and the discharge goals are set higher. In traumatic brain injuries, low cognitive scores translate into higher costs among those with higher motor function.

In summary, our analysis identified several potential areas of refinement for the payment system, assuming the analysis effects we observed hold up on 2003 data.

Acknowledgments

We thank the members of our Technical Expert Panel (TEP), whose names are listed below, for their support and guidance throughout this phase of our project. Members of the TEP reviewed an earlier version of this report; pointed out the need for additional analyses, which have been added here; and provided clinicians' views of the importance and significance of our findings. The report benefited greatly from the work of David Adamson, a RAND communications analyst, and Mark Totten, a RAND research programmer. Remaining problems of presentation and analysis are, of course, the responsibility of the authors.

Members of TEP on Inpatient Rehabilitation Facility–Prospective Payment System

Mindy Aisen Department of Veterans' Affairs	Sally Kaplan MedPAC
Ken Aitchison Kessler Institute for Rehabilitation	Richard Linn State University of New York Uniform Data System for Medical Rehabilitation
James Ball Catholic Health Services	John Melvin Jefferson Medical College of Thomas Jefferson University
Jean Davis HealthSouth Corp.	Paul Rao National Rehabilitation Hospital
Susan Dean-Baar University of Wisconsin	Pam Roberts Cedars Sinai Medical Center
Norbert Goldfield 3MHIS	Barry Smith Baylor Health Care Systems
Kurt Hoppe Mayo Clinic	Margaret Stineman University of Pennsylvania Medical Center
Robert Kane University of Minnesota	Carolyn Zollar AMRPA

1. Introduction

In 1999, RAND helped the Centers for Medicare and Medicaid Services (CMS; then known as the Health Care Financing Administration) develop a prospective payment system (PPS) for assigning costs to the treatment of specific conditions for Medicare patients receiving care at inpatient rehabilitation facilities (IRFs).

The PPS is based on assigning patients to particular function-related groups (FRGs) that attempt to predict the cost of treating particular Medicare patients according to how well they function in four general categories: transfers, sphincter control, self-care (e.g., grooming, eating), and locomotion.

The PPS is intended to align payments to hospitals as closely as possible with the actual costs of treating patients: If the government underpays for some kinds of care, hospitals have an incentive to limit access for patients requiring that kind of care; conversely, if the government overpays, resources are being wasted.

Conditions in the health care system change continually. Given that this newly implemented IRF PPS has been in place for over two years, CMS asked RAND to examine potential refinements to the FRGs to improve, where possible, the alignment between Medicare payments and actual hospital costs. There were several reasons to expect that such refinements would be possible:

- The data sets available are more complete: Instead of data on rehabilitation patients from just a sample of hospitals, we have new data, which describe the entire universe of rehabilitation patients.
- The data sets contain some new items: missing-value indicators and additional information about patient function status.
- We have implemented a number of improvements in the fitting algorithms that produced FRGs.
- Three years have passed, and changes have occurred in the cost structure of IRFs. Because such change is a constant in any PPS, it is likely that some periodic refitting of the FRGs will continue to be needed on an ongoing basis.

1.1. Purpose of This Report

This report presents our analysis of these potential refinements. We address two main questions:

(1) What new items are available in the data sets and how well do they improve the ability to predict costs?
(2) What refinements to the FRGs would allow better prediction of treatment costs?

To address the first question, we analyzed the new items from the data on IRF patient functioning and cost. When the IRF PPS was implemented in 2002, a new recording instrument was used to collect patient data. Known as the IRF Patient Assessment Instrument (or IRF PAI), the new instrument improved the quality of the information available to us.

To address the second question, we did two things: First, we considered whether alternative indices for patient functioning might predict costs more accurately; second, we ran the algorithm used in 1999 to derive FRGs with the new IRF PAI data to see whether the FRGs would look substantially different.

Existing FRGs quantify patient functioning according to 18 different measures. Two overall indices are used to summarize the 18 measures. These indices are known as FIM™, or functional independent measurement scores. The first of these is the "motor score," which sums up 13 of the 18 different components (we actually used 12; in our earlier work on developing FRGs, we excluded transfer to tub). The second of these is the "cognitive score," which sums up the remaining five components. In their current form, these indices do not weight any of the components they summarize. We used the data to test whether weighting would improve the indices' ability to predict costs.

1.2. Organization of This Report

The organization of this report is as follows:

In Section 2, we describe the new data and examine their potential for improving the prediction of Medicare rehabilitation treatment costs.

In Sections 3 and 4, we address the first of the main research questions noted earlier: What improvements are enabled by the new information?

In Section 3, we analyze the impact of the new information regarding "missing" data indicators analysis. The key question here is: Do the default missing values assigned in earlier FIM™ data seem appropriate?

In Section 4, we analyze the contribution of additional information about certain of the 18 FIM™ items to see if it helps to predict cost.

Sections 5 and 6 address the second of the main research questions: What refinements to the FRGs would improve prediction of treatment costs? In Section 5, we consider alternative indices to the standard motor and cognitive scores by exploring concepts of weighting.

In Section 6 we discuss the results of refitting FRGs to the new data, comparing new and old models and evaluating the improvement due to new indices.

Section 7 summarizes our conclusions regarding considerations for refinement.

2. Description of the New IRF PAI Data

In this section, we describe the earlier data used to create the PPS FRGs, identify the features of the new data, and highlight the new information made available in the new data set.

Specifically, there are three improvements in the data:

(1) In the earlier data set, "missing" information was simply coded as "needing maximal assistance." This coding may have been appropriate if the activity was too dangerous to perform, but not if the activity just happened to be unobserved or reflected a reliance on special equipment that may not have been available.

(2) New variables were added that provided further details on patient functioning. Standard bowel and bladder scores, for example, measure some combination of the level of assistance required and the frequency of accidents (i.e., soiling of clothes and surroundings); the new variables measure the level and the frequency separately.

(3) Additional optional information was recorded on the health status of patients—for example, shortness of breath, presence of ulcers, inability to balance. These were thought to be potentially useful predictors of cost.

We analyzed the power of all the new missing indicators and the patient functioning variables. The utility of the heath status group of variables was limited because many hospitals elected not to provide it; we chose not to analyze that here.

2.1. Number Of Cases

The new IRF PAI increased the sample size of our analytic file by almost 25 percent. Calendar year 2002 has 475,322 IRF cases, up from 390,048 in 1999, the data period on which the current system is based. The reasons for the increase are continued growth in the size of the patient population and an increase in the percentage of total patients in participating hospitals, from 64 to 100 percent. The larger file enables us to obtain greater precision in our estimates and ensures that we have a balanced and complete picture of the IRF PPS.

In our work based on 1999 data, we relied on information from three principal sources:

(1) Medicare program information (MEDPAR). The Medicare data files include discharge files recording patient-level demographic, clinical, and financial information, and hospital-level files providing facility characteristics and financial information.

(2) The Uniform Data System for medical rehabilitation (UDSmr™). UDSmr™ provided functional status and demographic information for rehabilitation discharges from participating hospitals.
(3) HealthSouth Corporation. HealthSouth provided similar functional status and demographic information for discharges from corporately owned rehabilitation hospitals during 1998 and 1999.

We linked hospital data from UDSmr™ and HealthSouth to MEDPAR to produce analysis files. The UDSmr™ and HealthSouth data gave us functional independence measures. The MEDPAR data told us about resource utilization: in particular, the charges levied by each hospital department for the particular inpatient-rehab stay. These charges were adjusted by cost-to-charge ratios, then again by hospital wage indices, to produce estimated case costs (for a discussion of the cost assignment method, see Newhouse et al., "Predicting Hospital Accounting Costs," *Health Care Financing Review,* Vol. 11, No. 1, pp. 25–33). We then attempted to model these costs, beginning with linking of hospital data to Medicare data.

The new IRF PAI has both Medicare identifiers and exact birth dates, so it links quite easily to the inpatient billing data. We performed this link using identifier or birth date, provider number, admission date, and discharge date, which yielded linked records for 94 percent of inpatient Medicare cases going to rehab hospitals. Table 2.1 shows the number of cases with usable information for our two main types of analysis (evaluating the new information and refitting FRGs).

There were 447,231 cases in 2002. For our analyses, we required complete information, including costs, rehabilitation impairment categories (RICs; 21 clinical groupings of patients that identify the reason for the rehabilitation stay), and function scores. Incomplete information reduced sample sizes by about 3 percent. The additional reduction to 330,724 is explained by selecting cases that were "typical": returned to the community, between four days and a year's duration, and not statistical outliers. The restriction to typical cases is standard for the development of FRGs. Because much of what we do with the new information is to determine what it adds to prediction accuracy, we consider the value of the new information just for typical cases as well.

An implication of our larger sample sizes is that we have enough precision to expect consistency across RICs. An observed effect has more credibility if it is replicated across most or all of the 21 RICs. So, we can look at both the strength and the consistency of an effect to determine its importance and reliability.

Table 2.1
Number of Usable Cases for Two Types of Analysis, by Rehabilitation Impairment Category (RIC)

RIC	Evaluating New Info	Refitting FRGs	Total Cases
1-Stroke	47,597	52,637	78,956
2-Traumatic BI	3,350	3,800	5,877
3-Nontraumatic BI	5,577	6,187	9,253
4-Traumatic SCI	1,330	1,531	2,469
5-Nontraumatic SCI	10,288	11,298	15,494
6-Neurological	13,279	14,725	20,029
7-Hip fracture	33,813	37,965	53,901
8-Major joint repl	81,681	89,622	104,615
9-Other orthopedic	14,118	15,790	21,541
10-Amputation, LE	7,319	8,102	11,857
11-Amputation, other	743	815	1,200
12-Osteoarthritis	7,420	8,182	10,517
13-Other arthritis	3,225	3,568	4,672
14-Cardiac	16,582	18,211	25,160
15-Pulmonary	7,038	7,666	10,498
16-Pain syndrome	6,979	7,707	9,964
17-MMT, no BI, SCI	2,916	3,420	4,883
18-MMT, w/BI or SCI	457	585	885
19-Guillain-Barré	375	412	598
20-Miscellaneous	34,501	38,337	54,596
21-Burns	141	164	266
Total	298,729	330,724	447,231

Note: The new information analysis was based on an earlier version of the 2002 data, which were complete only through November 2002. BI = brain injury; SCI = spinal cord injury; LE = lower extremity; MMT = major multiple trauma.

2.2. New Information

The FIM™ instrument used in developing the current payment system measured functional independence, but, in two respects, the scales were not perfect for our purposes. First, on the original FIM™ instrument, missing information was potentially confounded with low functional independence. Also, certain scores reflected multiple topics: bowel and bladder functioning were the minimum of level of assistance and frequency of accidents, which were not recorded separately in the FIM™ data; they are recorded separately here. We list in Table 2.2 the specific variables that provided new information.

In all cases, status is measured both at admission and at discharge; we use only the status at admission. The asterisked items had a special score for "missing," which enabled us to disentangle the meaning of "missing" from a default assignment.

The current analysis considers only the function modifiers and FIM™ instrument variables in the top part of the table (39R and above). Analysis of the other questions is complicated because the items are optional and, in our initial data sets, very often not filled in. We will monitor future releases of IRF PAI data to see whether these data begin to be recorded reliably.

Table 2.2
Questions in IRF PAI with Potentially Valuable New Information

QuestionType	Q#	Question
Function Modifiers	29	Bladder Level of Assistance
	30	Bladder Frequency of Accidents
	31	Bowel Level of Assistance
	32	Bowel Frequency of Accidents
	33	Tub Transfer*
	34	Shower Transfer*
	35	Distance Walked*
	36	Distance Traveled in Wheelchair*
FIM™ Instrument	37	Walk*
	38	Wheelchair*
	39A	Eating*
	39B	Grooming*
	39C	Bathing*
	39D	Dressing upper*
	39E	Dressing lower*
	39F	Toileting*
	39G	Bladder*
	39H	Bowel*
	39I	Transfers: bed, chair, wheelchair*
	39J	Transfers toil*
	39K	Transfers tub, shower*
	39L	Walk/wheelchair*
	39M	Stairs*
	39N	Comprehension*
	39O	Expression*
	39P	Social interaction*
	39Q	Problem solving*
	39R	Memory*
Medical Needs	25	Is patient comatose?
	26	Is patient delirious?
	27	Swallowing status
	28	Clinical signs of dehydration
Quality Indicators	48	Shortness of breath with exertion
	49	Shortness of breath at rest
	50	Weak cough and difficulty clearing
	51	Pain level
	52A	Pressure ulcer highest stage
	52B	Number of pressure ulcers
	52C	Surface wound area
	52D	Exudate amount
	52E	Tissue type
	52F	Total push score
	53	Balance problem

(*) missing items coded as zero. In data sets prior to the IRF PAI, the FIM™ instrument variables are coded as 1.0: maximal dependence.

3. Analysis of Missing FIM™ Data

This section and the next examine the new information made available by the IRF PAI data

The earlier FIM™ data did not distinguish between a response of "maximal dependence" and a response of "activity did not occur." The FIM™ measured functional independence in levels from 1.0 (maximal dependence) to 7.0 (minimal dependence); however, cases for which the activity did not occur were also coded as a 1.0. The IRF PAI distinguishes between the two: The "activity did not occur" response is coded as 0.0.

The 1.0 score for "missing" would mean "maximally dependent" if the patient is unable to do the activity either because it is too hard or too dangerous. But an activity may be unobserved if the necessary assistive device is not available, or the device the patient has at home is not available at the IRF, or not enough time has elapsed to observe the activity, or simply by human error.

There are pros and cons for assignment of maximal dependence. Under the current system, hospitals are better off if they do not code something because assigning "missing" as 1.0 pays them more. On the other hand, if CMS assigns higher (less dependent on assistance) default values for missing, that might encourage doing an activity (e.g., climbing stairs) when there is some risk of injury to the patient.

We investigate here the best default values for missing variables. We begin by characterizing the frequency of missing values of each function score. We then discuss the models we employed to estimate missing-value effects, and give the results of that estimation.

3.1. How Often Are FIM™ Items Missing?

Table 3.1 summarizes the missing-value status of the 18 FIM™ items. It is evident that the five cognitive variables are never missing (except for 29 cases in which data problems occurred), nor are bowel and bladder ever missing: This is consistent with instructions in the IRF PAI manual. We can see further that the top eight items are missing 1 percent or more of the time (rounded). The last three items are not missing very frequently, but perhaps might be a source of future concern to CMS.

Table 3.1
Frequency of FIM™ Scores, Present and Missing

Type of Item	Item	Number Present	Number Missing	% Missing
Motor	stairs	108,964	192,080	63.8
	transfer to tub	173,665	127,379	42.3
	bathing	288,414	12,630	4.2
	walking	290,146	10,898	3.6
	transfer to toilet	292,633	8,411	2.8
	dressing lower	298,054	2,990	1.0
	toilet	298,186	2,858	0.9
	dressing upper	298,660	2,384	0.8
	transfer to bed	299,978	1,066	0.4
	grooming	300,051	993	0.3
	eating	300,361	683	0.2
	bowel	301,044	0	0.0
	bladder	301,044	0	0.0
Cognitive	comprehension	301,015	29	0.0
	expression	301,015	29	0.0
	social interaction	301,044	0	0.0
	problem solving	301,044	0	0.0
	memory	301,044	0	0.0

Coding stairs "missing" as maximally dependent may seem like a good thing to do. Patients who are unstable should not be trying stairs: They may be too dangerous.

Transfer to tub has a high fraction of missing values, which we believe is due to a variety of causes. There are times when it is too dangerous to do the transfer. But there are also times when it is not done because the hospital lacks assistive devices. Empirical results support the equivocal nature of missing transfer to tub. Whenever we would examine numerical effects, transfer to tub alone seemed to consistently have the wrong sign (stronger performers cost more). That wrong sign could be caused by the 42 percent rating of cases as "maximally dependent" that in fact were not, so they were actually costing about what the mid-level transfer-to-tub performers were costing.

3.2. Models For Estimating Missing-Value Effects

The goal in this section is to develop equations that describe the relationship between costs and FIM™ scores. Right-hand-side variables to explain cost will include the standard measures: age, motor score, cognitive score, and comorbidities. In addition, we include dummy variables here to measure the missing-value effects. Our goal is to measure the (percentage) effect on cost of knowing that an item was actually missing versus scored as "maximally dependent."

In developing the FRGs, we had fit generalized additive models (GAM models) to the data. GAM is an exploratory tool that is particularly good at detecting nonlinear patterns in the data. GAM approximates a regression relationship as a sum of smooth (rather than linear) functions of the independent variables. As parameterized here, GAM divides the range of each variable into five equally spaced intervals and fits flexible curvers (cubic splines) within each interval. The method is described in Hastie et al., *The Elements of Statistical Learning,* New York, N.Y.: Springer-Verlag, 2001.

We continue to use GAM here to explore the relationships between candidate explanatory variables and cost. The basic form of our GAM model is

$$\text{logcost} = f_1(\text{motor}) + f_2(\text{cog}) + f_3(\text{age}) + \gamma\{\text{tier}\} \qquad (3.1)$$

where f_1, f_2, and f_3 are cubic splines fit by GAM and {tier} indicates comorbidity conditions. Here, as elsewhere in this document, we implement the tier effect as three dummy variables, tier-1 through tier-3, going from high to low levels of severity. The lists of tier diagnoses used here were developed from preliminary analysis of the same 2002 data. They are found in Carter and Totten (forthcoming) (although the definition of *ventilator condition* used there was not used here).

Because GAM approximates the relationship as a nonlinear function, a change in motor score from 20 to 21 might decrease predicted cost by a different percentage than might a change from 70 to 71. Because the relationship is assumed additive in the log scale, the decrease in predicted cost due to a change in motor score from 20 to 21 will be the same regardless of the values of the other independent variables. The flexibility that GAM affords can make its predictions more accurate than those of the linear model when the relationship between the dependent and independent variables is nonlinear.

We wish to examine the effect of the standard explainers of cost: motor score, cognitive score, and age. In addition, we want to control for all elements that might be missing, so that we can see whether the assignment of 1.0 introduced a bias. For each variable with percentage "missing" 1.0 percent or more (after rounding), we fit GAM models similar to what had been developed before, but with additional dummy variables to capture the effect of whether or not a given variable is missing:

$$\text{logcost} = f_1(\text{motor}_{-i}) + f_2(\text{cog}) + f_3(\text{age}) + \gamma_1\{\text{tier}\} + \gamma_2\{\text{missing}_{-i}\} + \beta_1 * \text{motor}_i + \beta_2 * \text{missing}_i \qquad (3.2)$$

So, for example, to investigate the effect of "missing" transfer to toilet, we ran a transfer-to-toilet model in which that variable was eliminated from the motor-score index, the standard variables were included, all "missing" indicator variables were included, and transfer to toilet was entered as a linear term. In Eq. (3.2), f_1, f_2, and f_3 are GAM fits, {tier} denotes the three levels of comorbidity dummy variables, motor_{-i} is motor score excluding the i^{th} component, $\{\text{missing}_{-i}\}$ denotes missing-value dummy variables excluding the i^{th} component, motor_i is the i^{th} component (keeping 1.0 as the default), and missing_i indicates whether the i^{th} component is missing. In each case, β_1 and β_2 contain the information about the effect of missing values. Because we kept the default assignment of 1.0 for motor_i, β_2 is the effect of being missing versus having a 1.0, $\beta_1 + \beta_2$ is the effect of being missing versus having a 2.0, etc.

These methods will tell us how much of a percentage shift in costs is explained by coding an activity as "missing" rather than "maximally dependent"; they also tell what score could be assigned to the missing activity other than the current default (1.0) to achieve the same effect as using the dummy variables.

3.3. Estimating Missing-Value Effects

We thought that the best way to summarize the information about missing-value effects was to solve for "neutral" missing values—observed $motor_i$ values for which predicted logcost is the same as for unobserved cases (in the current formulation, where $motor_i$ is assigned a 1.0, it can be shown that these "neutral" values are $1+\beta_2/\beta_1$). We computed these values for each variable, along with their standard errors. We then looked for consistency of these suggested missing-value scores across RICs. Table 3.2 shows the estimated missing-value scores for items missing with frequency 1 percent or more (rounded). Appendix Table A.1 shows the supporting details.

The first thing to note is that the currently assigned 1.0 values for stairs and walking appear right. They are almost never significantly different from 1.0. When they are significant, there is no pattern for assigning a "best" value. So, it seems that most of the missing-value cases could be following the instruction "do not test if too dangerous"—i.e., the activities would require maximal assistance.

On the other hand, Table 3.2 suggests that transfer to tub should get a 4.0. Values there cluster in the middle of the scale. They are almost always significant and greater than 1.0.

Using the missing value of 4.0 for transfer to tub causes the transfer-to-tub regression coefficient to take on the right sign. Appendix Table A.2 shows the regression coefficients for the 21 RICs. The transfer-to-tub regression coefficients are approximately equal to the motor-score regression coefficient, suggesting that transfer to tub might just as well be in the index: One point for transfer to tub affects costs about the same as one point for the average FIM™ item.

We looked for other cases where a variable had regression coefficients that were large enough so that a point or two in the missing-value score might be important. We think transfer to toilet is a candidate to receive a 2.0. Its values cluster around 2, and they are significantly greater than 2.0 in 10 RICs (t-statistics of 2 or more). And one point on the transfer-to-toilet scale is worth a lot: There is about a 6-percent drop in cost for every point increase in the FIM™ score for transfer to toilet (transfer to toilet is much more influential than the average FIM™ item). Appendix Table A.3 shows the supporting details for these observations.

Table 3.2
Suggested Missing-Value Scores for Frequently Missing Items

Item	% Missing	Median Estimated Value	Number of RICs With t-stat>=2
stairs	63.8	1	3
transfer to tub	42.3	4	18
bathing	4.2	2	4
walking	3.6	1	0
transfer to toilet	2.8	2	10
dressing lower	1.0	2	2
toilet	0.9	3	9
dressing upper	0.8	7	3

Note: A t-statistic of 2.0 or more occurs less than 2.5 percent of the time if the null hypothesis (here, best missing-value score = 1.0) is true.

3.4. Conclusions

The preceding analysis suggests that some minor improvements may be possible in building an index that specifically adjusts for "missing" status. Two categories in particular are affected when we no longer equate "missing" with "lowest functioning." Transfer to tub deserves a neutral score, and we reintroduce this variable into our analyses below with a missing-value score of 4.0. Transfer to toilet is another variable that appears to be frequently missing for reasons other than being too dangerous. It appears that the estimated difference in cost between "missing" and "highest dependence" is high—about 6 percent or more.

4. Analysis of Function Modifiers

This section looks at the usefulness of a new set of variables recorded in the IRF PAI called "function modifiers," shown in Table 2.2. These function modifiers provide supplemental information for the standard FIM™ scores. In several cases, instructions for the FIM™ scores had required observing multiple items and collapsing their values into a single score: For example, taking the minimum of the bowel level of assistance and frequency of accidents. The IRF PAI identifies which variables have multiple-component values and provides space for these to be recorded. We analyze the contribution to predicting costs of these additional items, below.

4.1. Function Modifiers for Bladder and Bowel

Bowel and bladder are given here as both level of dependence and frequency of accidents. Because their questions are worded the same (except for substitution of the word *bladder* for *bowel*), we analyze them together. Here is the bladder instruction:

> 29. Bladder level of assistance (Score using FIM™ levels 1–7)
>
> 30. Bladder frequency of accidents (score as below)
>
> 7 No accidents
> 6 No accidents, uses device such as a catheter
> 5 One accident in the past 7 days
> 4 Two accidents in the past 7 days
> 3 Three accidents in the past 7 days
> 2 Four accidents in the past 7 days
> 1 Five or more accidents in the past 7 days
>
> Enter in item 39G (Bladder) the lower (more dependent) score
> from Items 29 and 30 above

We found, first, that the FIM™ score reflects level of assistance more than it does accidents. The minimum of the two scores equals level of assistance a high percentage of the time: 92 percent for bladder, 95 percent for bowel. This is partly because the frequency of accidents question has changed. One accident per week used to get a 4.0 in the old FIM™ instrument. One accident per week now gets a 5.0 in the IRF PAI, and some other responses are coded upward as well.

To estimate the potential contribution of adding accident frequency, we modeled costs using the missing-value effects model from Section 3, excluding the bladder or bowel FIM™ item from the motor score, and trying two types of models. The first had level of assistance linear plus an indicator of accidents; the second had level of assistance linear plus five level-of-accidents dummy variables (1, 2, 3, 4, or 5+ accidents per week). To see effects that might be real, we looked for consistency and statistical significance across RICs. Using the five accidents-per-week dummy variables produced positive but unstable effects—these coefficients were not measured very precisely. The indicators of accidents, however, were positive and significant for most RICs. Table 4.1 gives the basic results for both bladder and bowel. The detailed regression results appear in Appendix Table A.4.

Bladder accidents are associated with an increase in costs of 5 or more percent in about half the RICs. Bowel accidents add at least that much in most of the RICs, and often more than 10 percent. These are large effects. Similarly to comorbidities, accidents seem to be an important cost multiplier in some RICs.

The frequency-of-accidents variable gets close to, but does not accurately measure, incontinence status, which is what we would really like to have. Incontinence adds to the burden of care for family and nursing staff, affects the rehabilitation plan, and can be the biggest barrier to discharge home. It is difficult to know what to make of frequency of accidents. It could possibly include liquid spills, or it could simply be a measure of quality of care (e.g., the nurse is not getting to the patient in time).

Even with these shortcomings, the frequency-of-accidents measure can be used as a proxy for incontinence. An increase in the short-term payment system to reflect bowel and bladder accidents might be considered. We also feel that the developers of the IRF PAI should give further attention to defining and refining an incontinence measure.

Table 4.1
Summary of Regression Accident Effects

Item	Median Cost Effect	Number of RICs with	
		t-stat>=2	t-stat>=1
bladder	4%	14	19
bowel	5%	16	20

Note: A t-statistic of 2.0 or more occurs less than 2.5 percent of the time if the null hypothesis (here, no bladder or bowel effects) is true. A t-statistic of 1.0 or more occurs about 16 percent of the time if the null hypothesis is true. The fact that these low-probability events occur in most of the 21 RICs implies that these outcomes are highly significant.

4.2. Distance Walked

We wanted to see if we could find useful information in function modifiers for patients with walking status (“W”) at admission. Thus, in our analysis, we excluded patients who began with wheelchair status (“C”) or both (“B”), which comprised 7.8 and 3.7 percent of the population, respectively.

The standard FIM™ walking item originally combined two types of items: distance walked and the amount of assistance required to travel certain distances. Neither of these is measured separately; the FIM™ item simply records whether the two are above or below joint thresholds. The function modifiers provide, separately, the level of assistance as well as the distance walked. The distance walked instruction says to code items as follows:

3: 150 feet
2: 50-149 feet
1: less than 50 feet
0: activity does not occur

Table 4.2 shows the relationship between the two function modifiers. As one might expect, there is a strong correlation between the two responses.

However, there are some surprises. With walking score of 1, you would not expect the patient to walk more than 50 feet (3.3 percent of the cases), although it can occur with a lot of help; likewise, walking scores of 5, 6, or 7 are somewhat inconsistent with the distance 1-49 category (5.8, 6.3, and 2.6 percent of the cases). We decided to investigate whether these types of anomalies might help in explaining cost. Are patients in the upper-right corner (low FIM™ scores, but longer distance) better off than the FIM™ score would indicate, and are patients in the lower-left corner (high FIM™ scores, shorter distance) worse off? We created variables indicating greater distance than expected for the lower walking levels and less distance than expected for the higher walking levels, then we added these to our GAM models.

Figure 4.1 shows that distance alone does at least as well as the FIM™ score in predicting costs. The measure of performance we chose in summarizing goodness of fit is R^2—the percentage of variance explained. In comparing two models, we regard the one with higher R^2 as being at least as good as the other. When we introduce a new model and observe a higher R^2 in most of the 21 RICs, we infer that we have a better model than the original. In the present case, 17 of the 21 RICs have higher R^2's when using the distance measure as opposed to the FIM™.

Table 4.2
Relationship Between the FIM™ Walking Score and Distance Walked

F-Mod Walk	Distribution of Observed Cases		
	1–49	50–149	150+
1	96.7%	2.7%	.6%
2	13.7	85.2	1.0
3	46.9	21.2	31.9
4	19.2	21.7	59.1
5	5.8	15.5	78.7
6	6.3	8.7	85.0
7	2.6	4.3	93.1

Figure 4.1—Comparison of R^2 Across the 21 RICs

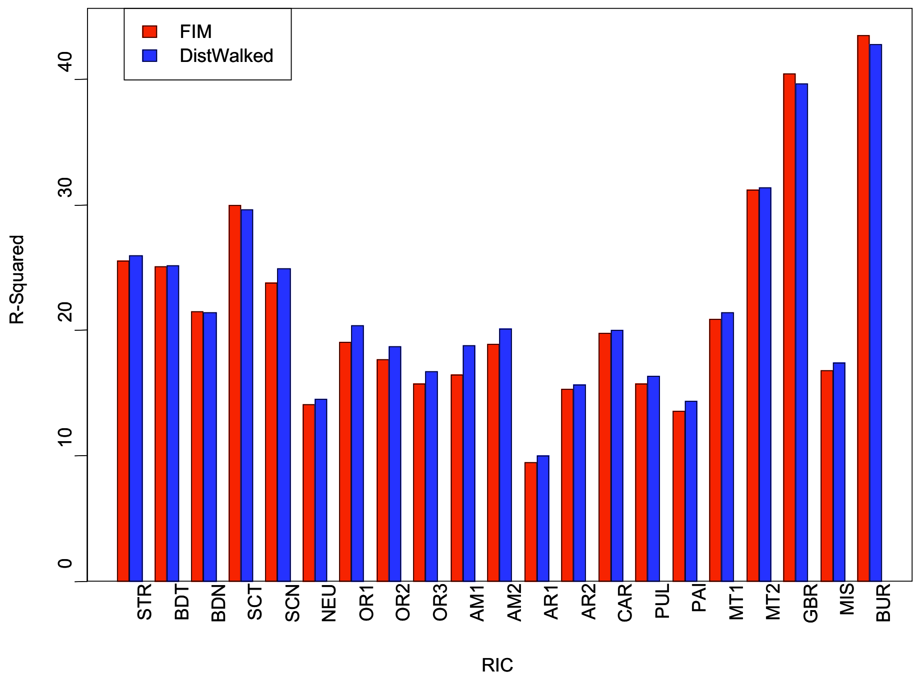


Table 4.3 shows the estimated distance effects—how performance on the distance measure translates to cost. This table summarizes a series of regression results contained in Appendix Table A.5.

Relative to patients who can walk 150 or more feet, the percentage increases in cost for patients who can only walk less than 150 or less than 50 feet are quite large and in the expected directions. Depending on the RIC, patients able to walk less than 50 feet cost 20 to 30 percent more than patients who can walk 150 or more feet, which is about twice the increase over those who can walk 50–149 feet. These effects are surprisingly large. They suggest that perhaps submerging distance walked in a composite FIM™ score, which itself is

Table 4.3
Increases in Cost Due to Decreased Walking Ability
(Relative to 150+-Feet Walkers)

Item	Median Cost Effect	Number of RICs with .t-stat>=2	Number of RICs with t-stat>=1
1–49	23%	21	21
50–149	10%	19	21
Missing	9%	17	20

Notes: A t-statistic of 2.0 or more occurs less than 2.5 percent of the time if the null hypothesis (here, no distance effects) is true. A t-statistic of 1.0 or more occurs about 16 percent of the time if the null hypothesis is true. The fact that these low probability events occur in most of the 21 RICs implies that these outcomes are highly significant.

one of only 12 components of a motor index, may not be the best thing to do for predicting cost.

Finally, we attempted to fit an interaction model to see whether distance walked added information above and beyond the FIM™ score. Our base model was again a GAM fit. We added to that interaction terms to indicate whether patients walked farther or shorter distance than expected, given their FIM™ scores. Those results are summarized in Table 4.4, abstracted from Appendix Table A.6.

The first of these interaction terms measures whether distance walked is less than the FIM™ scoring suggests. The second measures whether distance is more than the FIM™ suggests. It is the first of these that is highly significant. The percentage of variance explained goes up. Percentage effects are of a consistent sign, which, depending on the RIC, suggests cost increases in the range of 20 percent.

It is clear that distance walked has some additional explanatory power over the ordinary FIM™ score. But as currently worded, the distance variable would not be useful for payment purposes: It gives no indication of how much assistance was actually required for various distances. Measuring level of assistance would seem to be necessary: For example, how much assistance is required to achieve a fixed distance of 50 feet. It should also be determined whether assistance needed is determined by neurological as opposed to cardiac or pulmonary reasons.

Table 4.4
Results from a Distance-Walked/FIM™ Interaction Model

FIM™	Distance	Median Cost Effect	Number of RICs with t-stat>=2	Number of RICs with t-stat>=1
High	Low	19%	15	19
Low	High	1%	3	9

Note: A t-statistic of 2.0 or more occurs less than 2.5 percent of the time if the null hypothesis (here, no interaction effects) is true. A t-statistic of 1.0 or more occurs about 16 percent of the time if the null hypothesis is true. The fact that these low-probability events occur in most of the 21 RICs for the high-FIM™/low-distance entry implies that these outcomes are highly significant for this case.

4.3. Modifiers for Tub/Shower Transfers

The function-modifier instructions dictate scoring for both tub transfer and shower transfer and entering the lower of the two values as the FIM™ score. We found that one or the other is filled out, but not both. Thus, the only new information in the function modifiers is which of the two transfers is present.

We ran regression models with an indicator of the type of transfer environment: tub or shower. In regressions, patients measured as “tub” seem to cost more than patients measured as “shower”: Dummy-variable coefficients were positive in 19 of 21 RICs, with a median value of +3percent. It may be that “tub” is more costly than “shower” because a patient with a tub at home has to work harder to get ready for the tub, which is much more difficult and dangerous.

But we were not very enthusiastic about the finding. First, t-statistics were greater than 2.0 in only 11 of 21 cases, so the statistical effects did not hold up across a large fraction of the RICs. Second, transfer to tub was already a suspect variable because of its high frequency of "unobserved." Third, as we shall see in Section 5.2, even accounting for unobserved status, transfer to tub just does not show through as very important relative to the other FIM™ items, with or without the auxiliary information of whether the transfer is to shower or to tub. Hence, we decided not to investigate the utility of the tub/shower distinction further.

4.4. Conclusions

We would like to measure incontinence, but all we have from the current IRF PAI is frequency of accidents, for both bowel and bladder. We found that people who have accidents cost more, but we were unable to determine from the data available whether the higher-frequency accident patients cost more than the lower-frequency patients. We think that incontinence should be considered in a payment system much as comorbidities are treated now, and that before improving the IRF PAI so that it actually measures incontinence, the use of frequency of accidents as a proxy for incontinence should be considered.

Distance walked seems to contain as much or more information than the standard FIM™ walking measurement. We want to consider whether the FIM™ measurement can be improved so that its explanatory power is at least equal to distance walked. Some suggestions have been provided above.

5. Alternatives to the Standard Motor and Cognitive Scores

In this section and the next, we examine possible refinements to the FRGs to improve their prediction of costs.

This section examines the issue of weighting. Before we explore changes to FRGs, we assess whether it is possible to get motor and cognitive scores that better predict costs. We use what we call an "expanded" GAM method to explore alternative indices. This method allows us to assess the relative contribution of different kinds of impairment to treatment costs. Our goal is to obtain a weighting that is intuitively clear in what it measures and that brings about increases in predictive accuracy.

All of our past research utilized standard motor and cognitive scores, the sums of either 12 or 13 FIM™ motor items and the sum of 5 cognitive FIM™ items. This summing equally weights the items. While these indices have been accepted and used for many years, it is not clear whether they are optimal for predicting costs.

This section is organized as follows. First, we describe a method for modeling the cost data that will be used to develop candidate indices and to assess their predictive power. We then describe our application of this method to the 2002 data. Results include descriptions of average "optimal weights" and how much precision increase is associated with optimal weights. But because the optimal weights are a series of 378 points (21 RICs and 18 different weights), we cannot consider them in a potential payment-system context, so we explore some simplifications to produce candidate weightings, which we evaluate in Section 6.

5.1. Method for Developing Alternative Scores

In previous work, we explored the relationship of the FIM™ motor and cognitive scores to cost by fitting GAM models of the form

$$\text{logcost} = f_1(\text{motor}) + f_2(\text{cog}) + f_3(\text{age}) + \gamma\{\text{tier})\} \quad (5.1)$$

where f_1, f_2, and f_3 are GAM fits and {tier} denotes three levels of dummy variables.

Here, we explore fitting what we call expanded GAM models, of the form

$$\text{logcost} = f_1(\alpha_1{*}\text{eating} + \alpha_2{*}\text{grooming} + \ldots + \alpha_{13}{*}\text{stairs}) + f_2(\beta_1{*}\text{comprehension} + \ldots + \beta_5{*}\text{memory}) + \gamma_1\{\text{tier}\} + \gamma_2\{\text{age}\} \quad (5.2)$$

In fitting these models, we determine how to optimally weight the components of the motor and cognitive scores. The functions f_1 and f_2 are the same types that we fit using GAM (cubic splines) in Section 3. The term {age} here refers to three dummy variables: age<=70, age71-80, age>=81 (we chose not to fit GAM models to age to avoid additional computational complexity). To simplify interpretation, we constrain the α_js and β_js to sum to 100. To fit the model, we iteratively search for adjustments to the α_js and β_js that increase R^2, followed by refitting of the smooth terms f_1, f_2, and the age and tier dummy coefficients.

We fit these functions to the IRF PAI data. Doing so gave us two main types of output: first, the values of the optimal weights; second, an indication of how much variation could be explained by an optimal weighting of the individual items.

5.2. Optimal Weights

5.2.1. Values of Optimal Weights

The optimal weightings consist of many factors: for each of 21 RICs, there are 18 different weights, 378 all together.

We are interested in specifying how to weight the various FIM™ items to better predict costs. But we cannot use all 378 coefficients. Some of the 378 are based on small samples and are measured with a lot of noise. Perhaps more important, a complex weighting system that differs across all 21 RICs would be too cumbersome to manage. Everyone understands the current summary motor and cognitive scores. To produce an understandable weighting system, the coefficients would probably need to be the same across RICs.

We hoped that these weights would be fairly constant across RICs; to the extent that they are not, we would not be able to obtain a simple weighting scheme that achieved much of the gain in R^2 from optimal weighting. We looked at numerous plots to convince ourselves that, for the motor scores at least, items weighted heavily in one RIC were indeed weighted heavily in most others. We summarize what we observed in Table 5.1, which shows the motor-component averages across RICs within IRF PAI functional measurement groupings: self-care, sphincter, transfer, and locomotion. The weights are normalized to sum to 100.0.

Regarding the ability of these FIM™ items to predict cost, it appears that a subset of the self-care items could carry most of the self-care weight, that bowel and bladder might not be very important, that transfer to tub might again be rejected because its weight was considerably lower than the other transfer items, and that the transfer and locomotion weights might be a somewhat larger than the others.

We also computed averages of optimal weights for the cognitive items, as shown in Table 5.2 (again, the weights are normalized to sum to 100.0).

Table 5.1
Optimal Weights, Averaged Across RICs: Motor Items

Item Type	Functional Independence Item	Average Optimal Weight
Self	dressing lower	11.6
Self	toilet	8.4
Self	bathing	5.6
Self	eating	5.3
Self	dressing upper	2.5
Self	grooming	1.7
Sphincter	bladder	4.6
Sphincter	bowel	1.0
Transfer	transfer to bed	19.0
Transfer	transfer to toilet	10.6
Transfer	transfer to tub	5.9
Locomotion	walking	13.1
Locomotion	stairs	10.7
	total	100.0

Table 5.2
Optimal Weights, Averaged Across RICs: Cognitive Items

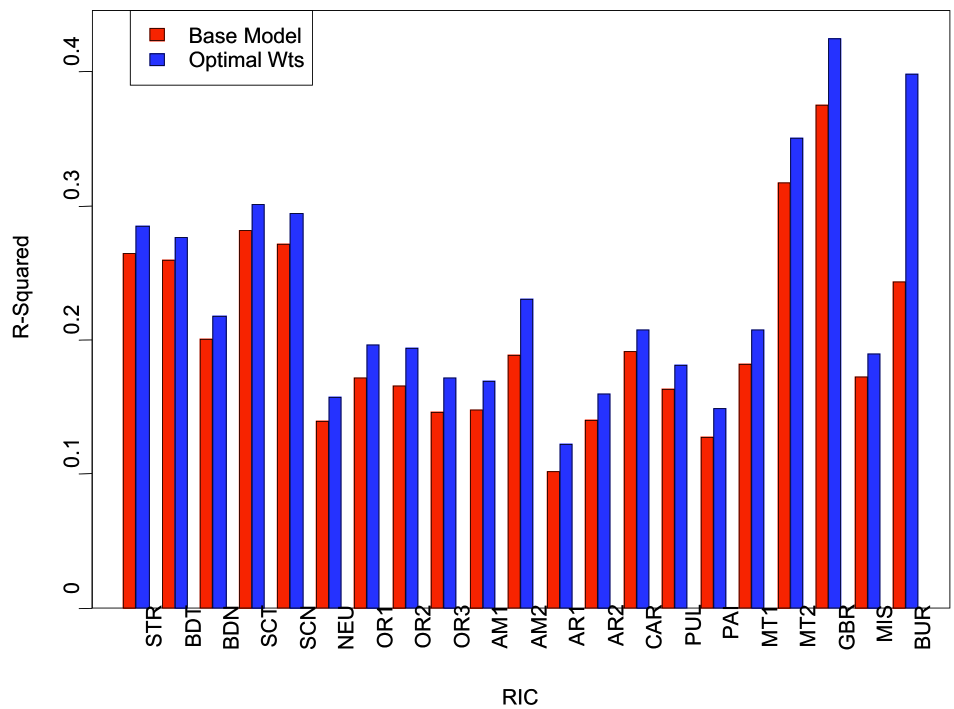
Item Type	Functional Independence Item	Average Optimal Weight
Cognitive	problem solving	27.7
Cognitive	memory	24.3
Cognitive	comprehension	21.0
Cognitive	social interaction	18.8
Cognitive	express	8.2
	total	100.0

But unlike motor, the cognitive items did not display a consistent pattern across RICs. The weights seemed to jump around almost randomly. Given this observation, and in view of the fact that the cognitive score had never been as important as the motor score in predicting cost, we did not pursue obtaining a better weighting of the cognitive indices.

5.2.2. Gain in Model Quality Due to Optimal Weights

Figure 5.1 shows the potential gain in R^2 (our chosen measure of model quality) from optimally weighting the components of the FIM™ scores. Here, we use the full set of 378 weights, 18 for each of 21 RICs. We compare the R^2 from these fits to the R^2 we would get from fitting our basic model with standard motor and cognitive scores (motor score excludes transfer to tub). As the graphic shows, we can always get some improvement in R^2 when we optimally weight the components of the motor and cognitive index. Some of these improvements are quite large, especially in smaller RICs such as 11 and 21, in which the number of parameters fit is a nontrivial fraction of the number of cases.

Figure 5.1—Weighting the FIM™ Items Differentially: Potential Gain in R^2



5.3. Simple Indices Based on Optimal Weights

We chose to stay with the standard cognitive index, but selected some simple motor indices (i.e., constrain the motor weightings) to see how much of the variance for optimally weighted items would be explained in fitting expanded GAM models.

At this point, we reintroduced into the analysis the function modifiers for sphincter and bowel. The optimal weights on bowel and bladder had been so low (Table 5.1) that we decided to give bowel and bladder every possible chance to show through as predictors of costs. We entered them into the analyses in their raw form.

We saw index selection as proceeding in two steps. First, we attempted to identify key components of the four sub-indices. For simplicity, we wanted these components to be equally weighted yet still explain a high fraction of the potential R^2 improvement. Second, we attempted to use simple linear combinations of the four sub-indices (rounded weights), again seeking to explain a high fraction of the potential R^2 improvement.

5.3.1. Building Sub-Indices That Perform Well

We first identified sets of variables within sub-index groups that had similar average weights. Table 5.3 shows the groupings we explored. For example, we tried an index Self1, which was simply

$$\text{Self1} = \text{dresslo} + \text{toilet} + \text{bathing} + \text{eating} \qquad (5.3)$$

Table 5.3
Optimal Weights and Candidate Index Types: Motor Items

Sub-Index Type	FIM™ Item	Average Weight	Index Type A	Index Type B
Self	dressing lower	9.9	Self1	Self2
Self	toilet	9.4	Self1	Self2
Self	bathing	7.3	Self1	Self2
Self	eating	5.9	Self1	
Self	dressing upper	2.6		
Self	grooming	1.3		
Sphincter	bladder-level	3.7	Sph3	
Sphincter	bladder-frequency		Sph3	Sph4
Sphincter	bowel-level	0.5	Sph3	
Sphincter	bowel-frequency		Sph3	Sph4
Transfer	transfer to bed	18.6	Xfr5	MobLoc7
Transfer	transfer to toilet	12.4	Xfr5	MobLoc7
Transfer	transfer to tub	4.3		
Locomotion	walking	13.4	Loc6	MobLoc7
Locomotion	stairs	10.8	Loc6	MobLoc7

We thought it would capture the effect of the self-care items because their optimal weights were about equal and somewhat higher than the remaining two self-care items. As another illustration, to capture mobility and transfer effects, we considered

$$\text{MobLoc7} = \text{trfbed} + \text{trftoil} + \text{walking} + \text{stairs} \tag{5.4}$$

In each case, the index-grouping column defines a sub-index equal to the sum of its components.

We tried several combinations to see how much of the variance for optimally weighted items they would explain and what an optimal weighting of the indices would be. We assumed that any alternative to the standard equally weighted motor index would have to be constant across RICs. So, we tried several candidate weightings to see how well they performed. We saw that Self1 produced better fits than Self2. We tried weightings on Xfr5 (in table) in the vicinity of twice those of Self1 and about 1.5 times those of Loc6, and we experimented with sphincter variables weighted up to half of Self1.

Here are some very simple alternatives that performed well and that used increasing amounts of information:

$$\text{Alt1} = \text{Self1} + 2 \times \text{MobLoc7} \tag{5.5}$$

$$\text{Alt2} = \text{Self1} + 2 \times \text{Xfr5} + 1.5 \times \text{Loc6} \tag{5.6}$$

$$\text{Alt3} = \text{Self1} + 2 \times \text{Xfr5} + 1.5 \times \text{Loc6} + .5 \times \text{Sph3} \tag{5.7}$$

These alternatives gave us three combinations to investigate all together, but using them required some arbitrary decisions about what to include and with what multiplier. So we added a fourth method that grew out of an objective algorithm for obtaining optimal weights. It was derived through the following five steps:

(1) Use the variables as they appear in the payment system. Thus, drop transfer to tub (a variable with too many problems, including missing information and low explanatory power) and use the FIM™ bladder and bowel scores without adding in the function modifiers.
(2) Compute a new set of optimal motor weights (for each of 21 RICs, there are 12 different weights, or 252 all together).
(3) Average those weights across RIC, and rescale the weights so that their average value is 1.0. This step is included so that the motor scores will range between 12 and 84, the range for the standard motor score, so the new motor score values will have a familiar meaning.
(4) Round to one decimal place to simplify.
(5) Designate a new index, Alt4, as this linear combination of the 12 motor-score components.

This alternative yields the motor-score index weightings shown in Table 5.4, which distributes its weight in a manner similar to the weights in Table 5.3. Note that transfers to bed or toilet get more of the weight than do any other variables, consistent with the finding of Stineman et al. that transfers were the strongest predictors of length of stay (Stineman et al., "Four Methods for Characterizing Disability in the Formation of FRGs," *Archives of Physical Medicine and Rehabilitation,* Vol. 75, December 1994, pp. 1277–1283).

Table 5.4
Definition of Optimal Weight Index Alt4: Motor Items

Sub-Index Type	FIM™ Item	Alt4 Index Wgts
Self	dressing lower	0.8
Self	toilet	1.0
Self	bathing	0.8
Self	eating	0.7
Self	dressing upper	0.3
Self	grooming	0.4
Sphincter	bladder	0.4
Sphincter	bowel	0.1
Transfer	transfer to bed	2.3
Transfer	transfer to toilet	1.8
Transfer	transfer to tub	0.0
Locomotion	walking	1.7
Locomotion	stairs	1.3

5.3.2. Performance of the Simple Indices

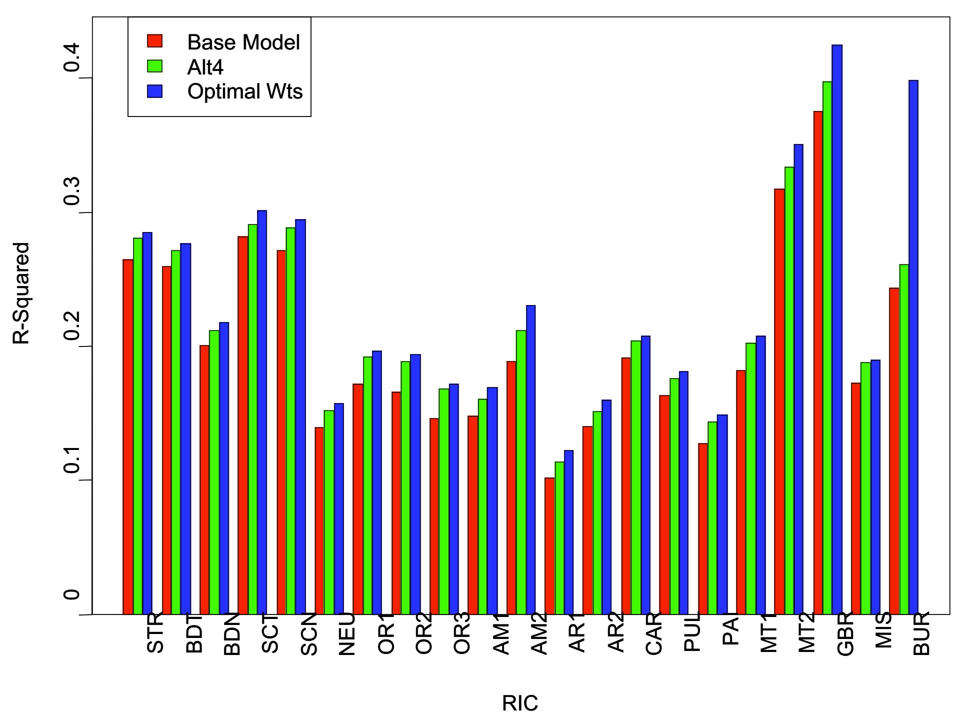
We fit costs using GAM separately for each of the four indices identified above and looked at the percentage of R^2 achieved between the base model and optimal weighting models. These numbers are summarized in Table 5.5 and appear in Appendix Table A.7. The simpler indices Alt1, Alt2, Alt3 explain less of the R^2 than does Alt4. One would expect such divergence because the weights of the first three alternatives bear less relation to the optimal weights. Among the simpler indices, the more information used, the better the index performed. The optimal weighting index Alt4 does considerably better than Alt1 through Alt3.

The main finding is that, except for burns (a very small RIC), these percentages are quite high and suggest that any one of these indices might effect significant improvement over the simple motor index in predicting costs. Figure 5.2 shows what fraction of the potential gain in R^2 (between the base model, which uses the 12-component motor score, and the RIC-by-RIC optimal weighting) is achieved by using Alt4. Note especially the staircase effect within each triple, with the middle bar about two-thirds of the distance between the leftmost and rightmost bars. This indicates that the improvements are occurring in the majority of the RICs.

Table 5.5
Percentage of R^2 Gain Achieved by Simple Indices That Are Constant Across RICs

	Simple Indices			
OutcomeMeasure	Alt1	Alt2	Alt3	Alt4
Average	56.5	60.4	62.0	66.9

Figure 5.2—Assessing How Well Alt4 Achieves the Potential Gain in R^2 for GAM Models



5.4. Conclusions

The results of this section suggest that we might expect some improvement in explanatory power by using a motor index that does not equally weight all components. We adopted a fitting method that determines a set of optimal weights, and we used those weights to suggest alternative indices that are constant across RICs. We tried several candidates. All performed better than the 12-component equally weighted motor score in GAM models. Of course, these are just GAM fits, designed to give us some indication of what might be best for a developing payment groups. The ultimate test of a new index is how its associated payment groups are able to distinguish between different levels of cost. We explore this performance in the following section.

6. Updating and Refitting the FRGs

CMS is now paying hospitals based on FRGs developed from 1999 data. Two questions arise from that observation.

(1) Would the FRGs look the same if they were based on 2002 data?
(2) How well do the 1999-based payments reflect costs in the 2002 data (e.g., how much change in the relationship between costs and cost predictions has occurred between 1999 and 2002)?

We attempt to answer these questions here.

To answer question (1), we updated our FRGs using the standard motor and cognitive indices, and we compared the form of the new (2002) FRGs with that of the old (1999). To answer question (2), we updated the 1999 FRGs in three different ways and compared their predictive performances to one another. The three ways correspond to increasingly aggressive changes from the current system: recalibrating payment levels for the 95 FRGs, refitting FRGs using the standard motor and cognitive scores (i.e., the update we refer to in question [1]), and refitting FRGs using the new indices derived in Section 5.

6.1. Updating the FRGs

Using 1999 data, we earlier developed 95 FRGs, the result of using the CART algorithm (Breiman et al., *Classification and Regression Trees*, Belmont, Calif.: Wadsworth, 1984). The construction of FRGs using CART involves recursively partitioning patients within RICs according to their age and their motor and cognitive FIM™ scores. CART produces the partitions so that the reported wage-adjusted rehabilitation cost of the patients captured is relatively constant within as opposed to between partitions. After producing the trees, we used judgment to trim the number of nodes back, to enforce certain constraints on the payment system, and to fit comorbidity tiers. The use of judgment, rather than a computer algorithm, made it nearly impossible to explore variations in CART models satisfying payment-system constraints. Also, the sequential nature of these steps introduced the risk that nodes would be introduced into the tree because a splitting variable was essentially correlated with important comorbidity effects.

Here, we employ a new algorithm that does all of this splitting and trimming iteratively. At each step of the tree construction, the CART algorithm enumerates all possible splits on all available predictor variables. It selects the variable and the split point that offers the greatest decrease in prediction error as measured by mean-squared error. But it considers only splits that satisfy a series of constraints:

(1) Neighboring FRGs have to differ by at least $1,500, unless eliminating that split changed the estimated cost by more than $1,000.
(2) Estimated costs had to be "monotone," decreasing in the motor and cognitive scores. That is, if either score is higher, presumably the patient has higher functioning and should cost less.
(3) Nodes should not have fewer than 50 observations.

The only split candidates are those that satisfy all the constraints and, therefore, the final set of FRGs are guaranteed to adhere to them as well.

6.1.1. Creating FRGs from 2002 Data

We ran the new algorithm on the 2002 data. Table 6.1 shows the FRGs that resulted. The following table identifies 95 groups, first by RIC, then by the function-related characteristics that define each group. The first entry, FRG 0101, has patients from the stroke RIC (01) with high motor scores (>55.5), and the last has patients from the burns RIC (21) with low motor scores (<38.5).

It is common to refer to these lists of groups as "classification trees," to alternately refer to classification groups as "tree nodes," and to define the tree by the criteria for being in each group.

Overall, our 2002 runs produced many fewer nodes—83instead of the 95 before—so the trees and, hence, the payment categories look somewhat different. The main differences were in stroke, which had 14 nodes before but only 8 now. In most of the RICs, costs are driven simply by motor score. The exceptions are stroke and traumatic spinal cord injury, in which younger people cost more, probably because they can take more therapy and their discharge goals are set higher. In traumatic brain injuries, low cognitive scores cost more among those with higher motor function.

6.1.2. Comparison of 2002 FRGs with 1999 FRGs

Table 6.2 shows the FRGs upon which the payment system were based, to be compared to the newly fit FRGs in Table 6.1. We were initially concerned that these FRGs looked different from the FRGs used in the payment system: One would prefer them to be stable so that the associated payment system would not have to change much over time. The reduction in number of nodes from 95 to 83 and in stroke from 14 to 8 has already been noted. Cognitive scores barely come through—only for RIC 2 for 2002, whereas they are present in RICs 1, 2, 5, 8, 12, and 18 for 1999. We note that the effect of age is about the same in the two sets of FRGs: Age appeared in RICs 1 and 4 in 1999, and in RICs 1 and 20 in 2002.

Table 6.1
FRGs Produced by the 2002 Data

RIC	FRG	Condition	N	Average Cost ($)
1-Str	0101	mot>55.5	4,089	7,504
1-Str	0102	mot>49.5 & mot<55.5	6,973	9,194
1-Str	0103	mot>46.5 & mot<49.5	4,520	10,318
1-Str	0104	mot>40.5 & mot<46.5	8,943	11,946
1-Str	0105	mot>32.5 & mot<40.5	10,707	14,684
1-Str	0106	mot<32.5 & age>81.5	4,552	17,116
1-Str	0107	mot<32.5 & age<81.5 & mot>26.5	4,861	18,001
1-Str	0108	mot<32.5 & age<81.5 & mot>26.5	7,648	20,584
2-Bdt	0201	mot>50.5 & cog>31.5	158	6,170
2-Bdt	0202	mot>50.5 & cog<31.5	687	8,799
2-Bdt	0203	mot>37.5 & mot<50.5 & cog>20.5	900	10,101
2-Bdt	0204	mot>37.5 & mot<50.5 & cog<20.5	533	12,344
2-Bdt	0205	mot>28.5 & mot<37.5	682	13,729
2-Bdt	0206	mot<28.5	817	17,862
3-Bdn	0301	mot>53.5	788	8,129
3-Bdn	0302	mot>40.5 & mot<53.5	2,326	10,136
3-Bdn	0303	mot>29.5 & mot<40.5	1,659	12,920
3-Bdn	0304	mot>16.5 & mot<29.5	1,092	15,743
3-Bdn	0305	mot<16.5	281	20,246
4-Sct	0401	mot>41.5	485	8,953
4-Sct	0402	mot>31.5 & mot<41.5	358	12,744
4-Sct	0403	mot>20.5 & mot<31.5	372	18,244
4-Sct	0404	mot<20.5 & age>57.5	204	25,909
4-Sct	0405	mot<20.5 & age<57.5	108	19,122
5-Scn	0501	mot>50.5	2,469	6,705
5-Scn	0502	mot>40.5 & mot<50.5	3,946	8,584
5-Scn	0503	mot>38.5 & mot<40.5	705	10,210
5-Scn	0504	mot>28.5 & mot<38.5	2,571	12,450
5-Scn	0505	mot<28.5	1,545	17,850
9-Or3	0901	mot>53.5	1,516	6,473
9-Or3	0902	mot>41.5 & mot<53.5	6,936	8,063
9-Or3	0903	mot>33.5 & mot<41.5	3,971	9,742
9-Or3	0904	mot<33.5	3,232	11,596
10-Am1	1001	mot>55.5	813	8,381
10-Am1	1002	mot>43.5 & mot<55.5	3,074	11,052
10-Am1	1003	mot<43.5	4,136	14,400
11-Am2	1101	mot>42.5	459	9,042
11-Am2	1102	mot<42.5	345	12,878
12-Ar1	1201	mot>46.5	3,473	8,291
12-Ar1	1202	mot>37.5 & mot<46.5	2,627	9,960
12-Ar1	1203	mot<37.5	2,017	11,908
13-Ar2	1301	mot>50.5	843	7,350
13-Ar2	1302	mot>40.5 & mot<50.5	1,299	9,083
13-Ar2	1303	mot<40.5	1,399	11,669
14-Car	1401	mot>55.5	2,507	6,499
14-Car	1402	mot>45.5 & mot<55.5	6,775	8,088
14-Car	1403	mot>35.5 & mot<45.5	5,597	9,888
14-Car	1404	mot<35.5	3,206	12,435
15-Pul	1501	mot>58.5	964	7,506
15-Pul	1502	mot>46.5 & mot<58.5	3,193	8,929
15-Pul	1503	mot>32.5 & mot<46.5	2,551	11,138
15-Pul	1504	mot<32.5	899	14,293
16-Pai	1601	mot>44.5	4,121	7,212
16-Pai	1602	mot>38.5 & mot<44.5	1,648	8,699
16-Pai	1603	mot<38.5	1,889	10,503

Table 6.1—Continued

RIC	FRG	Condition	N	Average Cost ($)	RIC	FRG	Condition	N	Average Cost ($)
6-Neu	0601	mot>49.5	3,424	8,113	17-Mt1	1701	mot>44.5	941	8,487
6-Neu	0602	mot>41.5 & mot<49.5	3,951	9,825	17-Mt1	1702	mot>31.5 & mot<44.5	1,443	11,392
6-Neu	0603	mot>35.5 & mot<41.5	2,533	11,042	17-Mt1	1703	mot<31.5	1,010	14,414
6-Neu	0604	mot<35.5	4,705	13,258					
					18-Mt2	1801	mot>33.5	310	10,707
7-Or1	0701	mot>50.5	4,072	7,450	18-Mt2	1802	mot>23.5 & mot<33.5	145	15,149
7-Or1	0702	mot>42.5 & mot<50.5	10,392	9,201	18-Mt2	1803	mot<23.5	125	22,563
7-Or1	0703	mot>33.5 & mot<42.5	12,673	11,144					
7-Or1	0704	mot<33.5	10,471	13,131	19-Gbr	1901	mot>41.5	171	9,583
					19-Gbr	1902	mot>32.5 & mot<41.5	90	16,403
8-Or2	0801	mot>57.5	5,441	5,217	19-Gbr	1903	mot<32.5	151	24,054
8-Or2	0802	mot>45.5 & mot<57.5	43,598	6,394					
8-Or2	0803	mot>36.5 & mot<45.5	27,897	7,777	20-Mis	2001	mot>53.5	5,474	7,042
8-Or2	0804	mot>30.5 & mot<36.5	7,869	9,255	20-Mis	2002	mot>41.5 & mot<53.5	15,833	8,871
8-Or2	0805	mot<30.5	3,912	10,897	20-Mis	2003	mot>32.5 & mot<41.5	9,448	10,815
					20-Mis	2004	mot<32.5	7,272	13,282
					21-Bur	2101	mot>38.5	87	12,449
					21-Bur	2102	mot<38.5	74	23,198

Table 6.2
FRGs Produced by the 1999 Data

RIC	FRG	Condition	N	Average Cost ($)	RIC	FRG	Condition	N	Average Cost ($)
1-Str	0101	mot>68.5 & cog>22.5	250	5,064	10-Am1	1001	mot>60.5	465	7,809
1-Str	0102	mot>58.5 & mot<68.5 & cog>22.5	1,596	6,400	10-Am1	1002	mot>51.5 & mot<60.5	1,412	9,377
1-Str	0103	mot>58.5 & cog<22.5	551	7,927	10-Am1	1003	mot>45.5 & mot<51.5	1,465	10,938
1-Str	0104	mot>52.5 & mot<58.5	4,131	8,168	10-Am1	1004	mot<45.5 & mot>38.5	1,319	12,532
1-Str	0105	mot>46.5 & mot<52.5	6,720	9,967	10-Am1	1005	mot<38.5	1,495	14,794
1-Str	0106	mot>41.5 & mot<46.5	5,753	11,932					
1-Str	0107	mot>38.5 & mot<41.5	3,102	13,616	11-Am2	1101	mot>51.5	407	7,793
1-Str	0108	mot>33.5 & mot<38.5 & age>82.5	987	13,739	11-Am2	1102	mot<51.5 & mot>37.5	580	10,711
1-Str	0109	mot>33.5 & mot<38.5 & age<82.5	3,620	15,756	11-Am2	1103	mot<37.5	217	14,300
1-Str	0110	mot<33.5 & age>88.5	584	14,750					
1-Str	0111	mot>26.5 & mot<33.5 & age>81.5 & age<88.5	1,003	16,252	12-Ar1	1201	mot>54.5 & cog>33.5	640	6,027
1-Str	0112	mot<26.5 & age>81.5 & age<88.5	1,065	18,546	12-Ar1	1202	mot>54.5 & cog<33.5	489	7,578
1-Str	0113	mot<33.5 & mot>26.5 & age<81.5	3,763	18,233	12-Ar1	1203	mot>47.5 & mot<54.5	1,643	8,358
1-Str	0114	mot<26.5 & age<81.5	4,215	20,869	12-Ar1	1204	mot<47.5 & mot>38.5	1,403	10,342
					12-Ar1	1205	mot<38.5	861	12,772
2-Bdt	0201	mot>51.5 & cog>23.5	320	7,137					
2-Bdt	0202	mot>39.5 & mot<51.5 & cog>23.5	303	9,858	13-Ar2	1301	mot>53.5	583	6,668
2-Bdt	0203	mot>39.5 & cog<23.5	602	11,522	13-Ar2	1302	mot>46.5 & mot<53.5	660	8,358
2-Bdt	0204	mot<39.5 & mot>29.5	400	14,101	13-Ar2	1303	mot<46.5 & mot>35.5	710	10,097
2-Bdt	0205	mot<29.5	428	19,149	13-Ar2	1304	mot<35.5	397	13,427
3-Bdn	0301	mot>50.5	1,053	8,168	14-Car	1401	mot>55.5	2,139	6,298
3-Bdn	0302	mot>40.5 & mot<50.5	1,164	10,754	14-Car	1402	mot>47.5 & mot<55.5	2,747	7,871
3-Bdn	0303	mot<40.5 & mot>24.5	1,099	14,429	14-Car	1403	mot<47.5 & mot>37.5	2,200	9,887
3-Bdn	0304	mot<24.5	442	20,333	14-Car	1404	mot<37.5	1,018	12,657
4-Sct	0401	mot>49.5	398	7,785	15-Pul	1501	mot>60.5	925	7,662
4-Sct	0402	mot>35.5 & mot<49.5	599	11,282	15-Pul	1502	mot>47.5 & mot<60.5	2,461	9,072
4-Sct	0403	mot<35.5 & mot>18.5	604	16,236	15-Pul	1503	mot<47.5 & mot>35.5	1,367	11,328
4-Sct	0404	mot<18.5	282	21,248	15-Pul	1504	mot<35.5	629	15,460

Table 6.2—Continued

RIC	FRG	Condition	N	Average Cost ($)	RIC	FRG	Condition	N	Average Cost ($)
5-Scn	0501	mot>50.5 & cog>29.5	1,498	6,317	16-Pai	1601	mot>44.5	2,035	7,201
5-Scn	0502	mot>50.5 & cog<29.5	243	7,951	16-Pai	1602	mot<44.5	958	9,789
5-Scn	0503	mot>40.5 & mot<50.5	2,034	8,630					
5-Scn	0504	mot<40.5 & mot>33.5	923	11,837	17-Mt1	1701	mot>45.5	596	8,136
5-Scn	0505	mot<33.5	1,139	16,882	17-Mt1	1702	mot<45.5 & mot>32.5	750	11,339
					17-Mt1	1703	mot<32.5	333	15,139
6-Neu	0601	mot>55.5	1,244	6,988					
6-Neu	0602	mot>46.5 & mot<55.5	2,486	8,911	18-Mt2	1801	mot>44.5 & cog>32.5	58	6,470
6-Neu	0603	mot<46.5 & mot>35.5	2,924	10,982	18-Mt2	1802	mot>44.5 & cog<32.5	103	9,927
6-Neu	0604	mot<35.5	2,221	13,373	18-Mt2	1803	mot<44.5 & mot>25.5	229	14,516
					18-Mt2	1804	mot<25.5	87	25,336
7-Or1	0701	mot>51.5	3,491	7,030					
7-Or1	0702	mot>45.5 & mot<51.5	5,139	8,656	19-Gbr	1901	mot>46.5	224	9,274
7-Or1	0703	mot<45.5 & mot>41.5	3,567	10,047	19-Gbr	1902	mot<46.5 & mot>30.5	245	16,916
7-Or1	0704	mot>37.5 & mot<41.5	3,033	11,059	19-Gbr	1903	mot<30.5	143	25,591
7-Or1	0705	mot<37.5	5,397	12,620					
					20-Mis	2001	mot>53.5	4,721	7,016
8-Or2	0801	mot>57.5	5,161	5,029	20-Mis	2002	mot>44.5 & mot<53.5	7,627	8,752
8-Or2	0802	mot>54.5 & mot<57.5	4,839	5,552	20-Mis	2003	mot<44.5 & mot>32.5	6,539	10,949
8-Or2	0803	mot>46.5 & mot<54.5	18,287	6,393	20-Mis	2004	mot<32.5 & age>81.5	854	12,173
8-Or2	0804	mot<46.5 & cog>31.5	9,323	7,708	20-Mis	2005	mot<32.5 & age<81.5	1,812	14,415
8-Or2	0805	mot<46.5 & mot>39.5 & cog<31.5	3,083	8,544					
8-Or2	0806	mot<39.5 & cog<31.5	2,734	10,625	21-Bur	2101	mot>45.5	94	9,284
					21-Bur	2102	mot<45.5	119	17,677
9-Or3	0901	mot>53.5	1,713	5,991					
9-Or3	0902	mot>46.5 & mot<53.5	2,905	7,692					
9-Or3	0903	mot<46.5 & mot>37.5	2,915	9,339					
9-Or3	0904	mot<37.5	1,777	11,920					

The previous two sets of trees differ for several reasons. First, CART trees developed from similar but not identical data sets can appear to be quite different, even though their predictions may be close. CART can be somewhat indifferent to certain splits, making an almost arbitrary choice, which dramatically affects the tree nodes that come after that choice. But it will ultimately converge to the true model. The only question is whether we have given it enough split opportunities to do so. Second, we have a new set of comorbidities, and the comorbidity effects are being determined simultaneously with the trees, not in series as before. Third, some elements of the stopping rules are fairly arbitrary, designed to produce trees with about 100 nodes in our earlier sample. There is no guarantee that they will produce 100 nodes in a different data set. Finally, the payment system has changed, and one would expect some differences in the cost structure.

More important than the actual form of the tree is whether the predictions are about the same. We attempt to answer this question next.

6.2. Comparing the Performance of Alternative FRGS

CMS will be facing the choice of whether to update FRGs by one of three methods. To get a feel for what to expect, we wish to approximate those methods here and compare their performance. The methods are:

- M1: Keep the definitions of the current FRGs, but recalibrate the weights. Since we are restricting our analysis here to cases discharged to the community, we estimate the effect of recalibration using a regression model that contains dummies for each FRG and for each combination of RIC and tier.
- M2: Develop new FRG definitions using the standard motor-score index, as in Table 6.1 above.
- M3: Develop new FRG definitions using one or more of the simple motor-score indices derived in the Section 6. We show these FRG definitions in Appendix B.

We evaluate these alternatives by comparing the values of R^2's to one another. To get some understanding of how these alternatives differ from a flexible fitting model, we compare their R^2's to a GAM model based on the standard motor-score index. Table 6.3 shows that, across RICs, the models have progressively more predictive power as you go from M1 to M3. Appendix table A.8 contains the details, by RIC.

We report both a total R^2 and an average R^2 here. The total is the usual measure: an indication of what percentage of total variance is explained by RIC and FRG. It is sensitive to the sizes of the different RICs, however. It does not convey that in many of the RICs the

Table 6.3
R^2 Summary Performance of Alternative FRGs

Type of Index	Model	R-Squared Summary Total	R-Squared Summary Average
Std Motor	M1	0.3009	0.1753
	M2	0.3012	0.1824
	GAM	0.3141	0.2026
New Motor	M3-Alt1	0.3132	0.1936
	M3-Alt2	0.3141	0.1958
	M3-Alt3	0.3141	0.1967
	M3-Alt4	0.3163	0.2002

Note: Alt1 through Alt4 are candidate simple indices expected to function well on the basis of the GAM model fits. They are described in Section 5.3.1.

explanatory power goes up, and often by a quite a bit. The average R^2 helps to summarize that rise in explanatory power.

The table suggests that refitting FRGs does better than simply recalibrating the 1999 FRGs, and that refitting with the alternative motor indices does better still, as we found in Section 5.3. Mobility variables seem to do a better job explaining cost than do the other FIM™ variables.

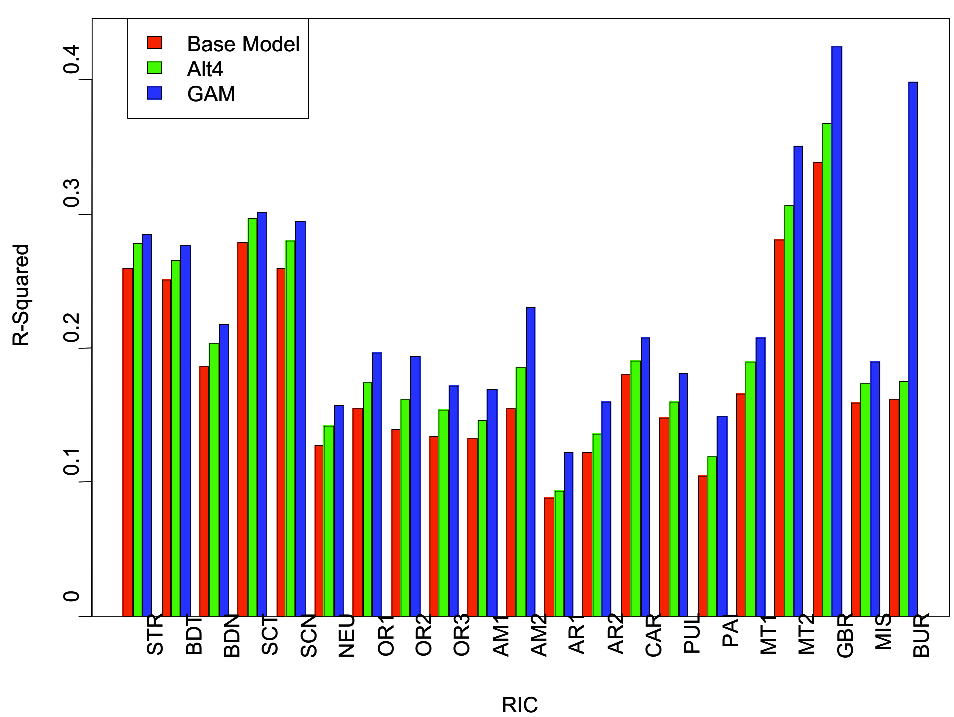
Figure 6.1 shows the fraction of potential gain in R^2 (between the base CART model, which uses the 12-component motor score, and the RIC-by-RIC optimal weighted GAM) achieved by using Alt4. Except for the very small burns RIC, it tends to cover more than half of the potential gain and, at times, almost all of it.

6.3. Conclusions

Updating FRGs with new data produces similar-looking, but different, trees—and one should expect some differences in any future update. Cut points may change or a new split may occur that almost but not quite made it into earlier runs, possibly leading to dramatically different trees.

Our update from 1999 to 2002 data led to only moderate changes in the trees. Cognitive scores and age played more of a role in the earlier FRGs; now, FRGs are mostly a function of motor scores. But, as we showed in Table 6.3, if we stay with the standard motor score, the explanatory power of the refit (2002) FRGs is about the same as the old (1999) FRGs.

Figure 6.1—Assessing How Well Alt4 Achieves the Potential Gain in R^2 for CART Models



7. Summary of Considerations for Refinement

Our analyses have identified several potential areas of refinement for the payment system. Assuming the analysis effects hold up on 2003 data, we recommend the following actions.

7.1. Missing Values

Concerning missing values: assigning transfer to tub a 4.0 seems well supported by the data. However, transfer to tub is not a good variable. It is too often missing. Even when present, it seems much less important than other motor-score components. The decision to drop transfer to tub from the motor score in developing the payment system was made partly because we did not have the information to break apart “missing” from “unable to perform.” We see no reason to reintroduce that variable now.

The transfer-to-toilet assignment of 2.0 seems, on the surface, less important than transfer to tub; however, transfer to toilet is a much more important explainer of cost. We think that the missing-value scoring on this variable should be changed to 2.0.

7.2. Function Modifiers

Bowel and bladder FIM™ items are currently in the FIM™ index, but their effect is not strong. However, the presence of bowel or bladder accidents seems to be an important cost multiplier in some RICs, similar to comorbidities. We think presence of accidents is mostly measuring incontinence, which is known to drive costs. But the IRF PAI ought to go after incontinence directly with questions eliciting the level of incontinence. Pending such a modification to the questionnaire, the presence of accidents might be considered as an adjustment.

Distance walked appears to be an important predictor variable. But the IRF PAI question needs to be refined so that it measures the burden of care required for the patient walk a set distance. In estimating the effect of distance, it seems advisable to consider what the limiting factors might be (e.g., neurological, pulmonary, cardiac).

7.3. Alternatives to the Standard Motor Index

Alternative motor indices predict better than the 12-component score used in developing the payment system. Numerous alternatives were considered, from simple indices based on judgment to an index produced by an objective optimal-weighting algorithm. The mobility items in all of these indices tended to be the most important predictors and received higher weight. Objective algorithms are easier to update and less controversial for the rehabilitation community to accept. In future payment-system updates, we think such an algorithm should be considered, along with the standard 12-component motor score; a decision on which one to use made should then be made, based on both the predictive power and clinical validity of the classifications. The downside of choosing a weighted index, of course, is complexity.

APPENDIX A

Supporting Tabular Details

In general, we require two conditions to be satisfied in identifying potentially important effects. First, in our model fits, we expect statistical significance. Second, we expect that a given result will replicate across a high number of the 21 RICs. The main body of the text summarizes our findings. The supporting detail generally requires extensive tables, with RIC defining the row. We have chosen to include summaries only in the text and to place the full tables in this appendix.

Table A.1
Optimal Missing-Value Scores for Frequently Missing Items

RIC	stairs	trftub	bathing	walking	trftoil	dressup	dresslo	toilet
1-Stroke	1	4 *	1	1	2 *	4 *	3 *	3 *
2-Traumatic BI	0	4 *	0	0	2 -	-8	-6	5 -
3-Nontraumatic BI	2 *	4 *	3 *	1	1	35	2	-5
4-Traumatic SCI	1	5 *	6	2	1	-28	-9	6 *
5-Nontraumatic SCI	0	4 *	3	1	1	35	4 -	0
6-Neurological	1	4 *	2	0	2 *	4	3 -	3 *
7-Hip fracture	0	4 *	2 *	1	2 *	8 -	1	3 *
8-Major joint replacement	1	4 *	2 *	1	2 *	-2	1	3 *
9-Other orthopedic	0	4 *	2 -	0	2 *	91	2 -	2 -
10-Amputation, LE	1	3 *	2 -	1	2 *	15 -	2 -	3 *
11-Amputation, other	3	6 -	1	12	3 *	10	0	1
12-Osteoarthritis	5 *	8 *	0	0	3 *	7 *	1	1
13-Other arthritis	11	6 *	2 -	0	3 *	8	4 -	2
14-Cardiac	0	6 *	1	1	0	32 -	0	4 -
15-Pulmonary	1	55	2 -	1	0	5 -	6 *	7 *
16-Pain syndrome	0	5 *	2	2	1	-9	-2	7 *
17-MMT, no BI, SCI	1	5 *	3 -	3	3 *	35	1	3 -
18-MMT, w/BI or SCI	3 -	3 -	25	2	0	14 -	33 -	13 -
19-Guillain-Barre	1	4 *	22	2	0	16	6 -	8 -
20-Miscellaneous	0	6 *	2 *	1	1	6 *	2 -	3 *
21-Burns	4 *	3 *	33	10	5	895	2	-4

Notes: (-) denotes whether the missing-value score is more than 1 standard error above 1.0; (*) denotes whether it is more than 2 standard errors above 1.0. BI = brain injury; SCI = spinal cord injury; LE = lower extremity; MMT = major multiple trauma.

Table A.2
Regression Coefficients When Transfer-to-Tub Missing Value Is Coded as 4.0

RIC	Coefficient	%-effect	motor-coef
1-Stroke	-0.021 *	-2	-0.029
2-Traumatic BI	-0.020 *	-2	-0.018
3-Nontraumatic BI	-0.021 *	-2	-0.018
4-Traumatic SCI	-0.017 -	-2	-0.029
5-Nontraumatic SCI	-0.032 *	-3	-0.030
6-Neurological	-0.024 *	-2	-0.019
7-Hip fracture	-0.024 *	-2	-0.021
8-Major joint replacement	-0.024 *	-2	-0.020
9-Other orthopedic	-0.015 *	-1	-0.020
10-Amputation, LE	-0.023 *	-2	-0.019
11-Amputation, other	-0.017 -	-2	-0.028
12-Osteoarthritis	-0.026 *	-3	-0.017
13-Other arthritis	-0.022 *	-2	-0.018
14-Cardiac	-0.020 *	-2	-0.021
15-Pulmonary	-0.006 -	-1	-0.019
16-Pain syndrome	-0.033 *	-3	-0.021
17-MMT, no BI, SCI	-0.018 *	-2	-0.020
18-MMT, w/BI or SCI	0.018	2	-0.022
19-Guillain-Barre	-0.048 *	-5	-0.043
20-Miscellaneous	-0.022 *	-2	-0.021
21-Burns	0.134	14	-0.040

Notes: (-) denotes whether the transfer-to-tub regression coefficient is more than 1 standard error below 0.0; (*) denotes whether it is more than 2 standard errors below 0.0. The %-effect column is the percentage change in cost per 1-point increase in transfer to tub. BI = brain injury; SCI = spinal cord injury; LE = lower extremity; MMT = major multiple trauma.

Table A.3
Regression Coefficients When Transfer-to-Toilet Missing Value Is Coded as 2.0

RIC	Coefficient	%-effect	motor coef
1-Stroke	-0.084 *	-8	-0.029
2-Traumatic BI	-0.066 *	-6	-0.018
3-Nontraumatic BI	-0.067 *	-6	-0.018
4-Traumatic SCI	-0.079 *	-8	-0.029
5-Nontraumatic SCI	-0.098 *	-9	-0.030
6-Neurological	-0.063 *	-6	-0.019
7-Hip fracture	-0.069 *	-7	-0.021
8-Major joint replacement	-0.059 *	-6	-0.020
9-Other orthopedic	-0.061 *	-6	-0.020
10-Amputation, LE	-0.060 *	-6	-0.019
11-Amputation, other	-0.106 *	-10	-0.028
12-Osteoarthritis	-0.053 *	-5	-0.017
13-Other arthritis	-0.061 *	-6	-0.018
14-Cardiac	-0.047 *	-5	-0.021
15-Pulmonary	-0.050 *	-5	-0.019
16-Pain syndrome	-0.060 *	-6	-0.021
17-MMT, no BI, SCI	-0.070 *	-7	-0.020
18-MMT, w/BI or SCI	-0.099 *	-9	-0.022
19-Guillain-Barre	-0.115 *	-11	-0.043
20-Miscellaneous	-0.057 *	-6	-0.021
21-Burns	0.028	3	-0.040

Notes: (-) denotes whether the transfer-to-toilet regression coefficient is more than 1 standard error below 0.0; (*) denotes whether it is more than 2 standard errors below 0.0. The %-effect column is the percentage change in cost per 1-point increase in transfer to toilet. BI = brain injury; SCI = spinal cord injury; LE = lower extremity; MMT = major multiple trauma.

Table A.4
Regression Effects of Incontinence on Cost

RIC	Bladder %-effect	Bowel %-effect
1-Stroke	2 *	2 *
2-Traumatic BI	2 -	11 *
3-Nontraumatic BI	2 -	3 *
4-Traumatic SCI	8 *	23 *
5-Nontraumatic SCI	5 *	10 *
6-Neurological	6 *	6 *
7-Hip fracture	2 *	5 *
8-Major joint replacement	6 *	6 *
9-Other orthopedic	5 *	5 *
10-Amputation, LE	2 -	6 *
11-Amputation, other	10 *	13 *
12-Osteoarthritis	8 *	8 *
13-Other arthritis	9 *	10 *
14-Cardiac	4 *	4 *
15-Pulmonary	1	1 -
16-Pain syndrome	3 *	4 *
17-MMT, no BI, SCI	3 -	2 -
18-MMT, w/BI or SCI	11 *	10 -
19-Guillain-Barre	4 -	4 -
20-Miscellaneous	3 *	4 *
21-Burns	0	-4

Notes: (-) denotes whether the incontinence regression coefficient is more than 1 standard error above 0.0; (*) denotes whether it is more than 2 standard errors above 0.0. BI = brain injury; SCI = spinal cord injury; LE = lower extremity; MMT = major multiple trauma.

Table A.5
Increases in Cost Due to Walking Ability
Relative to 150+-Feet Walkers)

	Percent Increases		
RIC	1-49	50-149	Missing
1-Stroke	25 *	10 *	16 *
2-Traumatic BI	19 *	9 *	10 *
3-Nontraumatic BI	16 *	3 *	11 *
4-Traumatic SCI	19 *	8 -	9 -
5-Nontraumatic SCI	35 *	13 *	17 *
6-Neurological	19 *	7 *	8 *
7-Hip fracture	29 *	11 *	9 *
8-Major joint replacement	23 *	9 *	6 *
9-Other orthopedic	25 *	10 *	8 *
10-Amputation, LE	46 *	20 *	7 *
11-Amputation, other	14 *	4 -	26 *
12-Osteoarthritis	17 *	8 *	11 *
13-Other arthritis	18 *	5 *	12 *
14-Cardiac	23 *	11 *	7 *
15-Pulmonary	23 *	10 *	5 -
16-Pain syndrome	22 *	9 *	13 *
17-MMT, no BI, SCI	23 *	12 *	21 *
18-MMT, w/BI or SCI	26 *	14 *	5
19-Guillain-Barre	31 *	23 *	7 -
20-Miscellaneous	23 *	9 *	7 *
21-Burns	34 *	45 *	73 *

Notes: Sample restricted to locomotion mode at admission = W. (-) denotes whether the estimated effect is more than 1 standard error above 0.0; (*) denotes whether it is more than 2 standard errors above 0.0. BI = brain injury; SCI = spinal cord injury; LE = lower extremity; MMT = major multiple trauma.

Table A.6
Results from a Distance-Walked / FIM™ Interaction Model

RIC	Percentage Increases in Cost: FIM™>=3 Dist<50	Percentage Increases in Cost: FIM™<=4 Dist>=150	R-Squared: Base Model	R-Squared: Interact. Model
1-Stroke	18 *	1 -	25.49	25.68
2-Traumatic BI	21 *	2 -	25.04	25.27
3-Nontraumatic BI	11 *	7 *	21.48	21.70
4-Traumatic SCI	9 -	5 -	29.96	30.05
5-Nontraumatic SCI	23 *	-1	23.77	24.19
6-Neurological	19 *	5 *	14.10	14.35
7-Hip fracture	26 *	0	19.06	19.69
8-Major joint replacement	19 *	0	17.68	18.08
9-Other orthopedic	22 *	0	15.77	16.19
10-Amputation, LE	29 *	-12	16.45	17.60
11-Amputation, other	17 -	3	18.89	19.15
12-Osteoarthritis	14 *	-1	9.47	9.71
13-Other arthritis	1	1	15.29	15.29
14-Cardiac	16 *	0	19.73	19.92
15-Pulmonary	19 *	-2	15.79	16.13
16-Pain syndrome	20 *	3 -	13.59	13.89
17-MMT, no BI, SCI	20 *	2 -	20.91	21.24
18-MMT, w/BI or SCI	10 -	-10	31.19	31.65
19-Guillain-Barre	0	10 -	40.42	40.57
20-Miscellaneous	17 *	2 *	16.82	17.03
21-Burns	29 -	-20	43.48	44.80

Notes: Sample restricted to locomotion mode at admission = W. (-) denotes whether the estimated effect is more than 1 standard error above 0.0; (*) denotes whether it is more than 2 standard errors above 0.0. BI = brain injury; SCI = spinal cord injury; LE = lower extremity; MMT = major multiple trauma.

Table A.7
Percentage of R^2 Gain Achieved by Simple Indices That Are Constant Across RICs

	Index Combinations			
RIC	Alt1	Alt2	Alt3	Alt4
1-Stroke	55.3	68.4	48.9	79.5
2-Traumatic BI	72.9	75.2	67.8	71.1
3-Nontraumatic BI	64.3	67.7	51.9	68.6
4-Traumatic SCI	11.4	20.4	85.1	48.5
5-Nontraumatic SCI	48.1	59.5	85.4	72.1
6-Neurological	59.8	68.4	79.1	72.7
7-Hip fracture	70.4	73.8	79.8	85.0
8-Major joint replacements	77.3	74.5	68.1	81.9
9-Other orthopedic	80.6	79.4	74.5	86.4
10-Amputation, LE	48.3	62.7	66.8	60.8
11-Amputation, other	48.0	55.1	48.7	56.1
12-Osteoarthritis	47.6	53.2	40.9	59.1
13-Other arthritis	59.5	58.8	52.2	59.6
14-Cardiac	70.5	67.8	72.6	80.4
15-Pulmonary	76.5	73.5	51.3	72.0
16-Pain syndrome	74.6	73.5	58.8	75.7
17-MMT, no BI, SCI	66.9	69.6	72.9	81.1
18-MMT, w/BI or SCI	40.9	43.1	66.5	48.8
19-Guillain-Barre	24.8	31.2	55.6	44.5
20-Miscellaneous	77.5	81.4	74.1	89.0
21-Burns	11.9	11.0	1.9	11.7
Average	56.5	60.4	62.0	66.9

Notes: BI = brain injury; SCI = spinal cord injury; LE = lower extremity; MMT = major multiple trauma.

Table A.8
R^2 Performance of Alternative FRGs

RIC	Standard Index Models			New Index Models			
	M1	M2	GAM	M3-Alt1	M3-Alt2	M3-Alt3	M3-Alt4
1-Stroke	0.2634	0.2594	0.2651	0.2764	0.2787	0.2744	0.2781
2-Traumatic BI	0.2382	0.2512	0.2597	0.2552	0.2527	0.2597	0.2660
3-Nontraumatic BI	0.1784	0.1866	0.2011	0.2042	0.2027	0.1999	0.2037
4-Traumatic SCI	0.2584	0.2791	0.2814	0.2891	0.2764	0.3007	0.2972
5-Nontraumatic SCI	0.2537	0.2593	0.2717	0.2684	0.2705	0.2821	0.2800
6-Neurological	0.1290	0.1273	0.1398	0.1384	0.1407	0.1421	0.1426
7-Hip fracture	0.1552	0.1552	0.1722	0.1693	0.1706	0.1715	0.1744
8-Major joint replacement	0.1377	0.1397	0.1659	0.1606	0.1594	0.1588	0.1618
9-Other orthopedic	0.1326	0.1345	0.1467	0.1494	0.1505	0.1529	0.1539
10-Amputation, LE	0.1396	0.1326	0.1483	0.1379	0.1409	0.1472	0.1467
11-Amputation, other	0.1556	0.1546	0.1887	0.1766	0.1814	0.1741	0.1857
12-Osteoarthritis	0.0947	0.0888	0.1020	0.0855	0.0945	0.0951	0.0937
13-Other arthritis	0.1317	0.1224	0.1400	0.1350	0.1318	0.1295	0.1363
14-Cardiac	0.1780	0.1802	0.1916	0.1860	0.1880	0.1886	0.1909
15-Pulmonary	0.1484	0.1484	0.1629	0.1580	0.1600	0.1501	0.1604
16-Pain syndrome	0.0907	0.1047	0.1273	0.1157	0.1157	0.1152	0.1192
17-MMT, no BI, SCI	0.1620	0.1659	0.1818	0.1812	0.1824	0.1902	0.1897
18-MMT, w/BI or SCI	0.2839	0.2809	0.3178	0.2614	0.2896	0.2950	0.3065
19-Guillain-Barre	0.3149	0.3385	0.3755	0.3622	0.3631	0.3517	0.3681
20-Miscellaneous	0.1616	0.1592	0.1728	0.1704	0.1717	0.1708	0.1734
21-Burns	0.0737	0.1616	0.2431	0.1848	0.1895	0.1815	0.1757
Total	0.3009	0.3012	0.3141	0.3132	0.3141	0.3141	0.3163
Average	0.1753	0.1824	0.2026	0.1936	0.1958	0.1967	0.2002

Notes: BI = brain injury; SCI = spinal cord injury; LE = lower extremity; MMT = major multiple trauma.

APPENDIX B

FRGs Based on Simple Indices

This appendix provides FRGs produced using the 2002 data and the alternative indices derived in Section 5. There are four tables, corresponding to Alt1 through Alt4. The FRG definitions are quite similar. The number of nodes ranges from 86 to 90. In most (16 or 17) of the RICs, costs are driven exclusively by motor score. Age is present in both stroke (1_Str) and spinal cord (4_Sct) injuries, for which younger people cost more: Presumably, they can take more therapy and goals for their recovery are higher. Age is also present in lower extremity joint replacement (8_Or2), in which younger people with high motor scores are costing less. Cognitive skills are present in the brain injury RICs (2_Bdt and 3_Bdn), but appear in stroke only for models Alt1 and Alt2.

Table B.1
FRGs Produced by the 2002 Data, Index Variation Alt1

RIC	FRG	Condition	N	Average Cost ($)	RIC	FRG	Condition	N	Average Cost ($)
1 Str	0101	mot>54.5	3,934	7,434	9 Or3	0901	mot>49.5	1,844	6,468
1 Str	0102	mot>48.5 & mot<54.5 & cog>23.5	3,574	8,491	9 Or3	0902	mot>39.5 & mot<49.5	5,742	7,944
1 Str	0103	mot>48.5 & mot<54.5 & cog<23.5	1,844	9,984	9 Or3	0903	mot>31.5 & mot<39.5	4,482	9,586
1 Str	0104	mot>42.5 & mot<48.5	8,895	10,396	9 Or3	0904	mot<31.5	3,587	11,617
1 Str	0105	mot>38.5 & mot<42.5	6,139	12,051					
1 Str	0106	mot>30.5 & mot<38.5	10,810	14,478	10 Am1	1001	mot>50.5	997	8,632
1 Str	0107	mot>25.5 & mot<30.5	5,761	17,665	10 Am1	1002	mot>40.5 & mot<50.5	2,455	10,789
1 Str	0108	mot<25.5 & age>80.5	3,425	18,147	10 Am1	1003	mot<40.5	4,571	14,292
1 Str	0109	mot<25.5 & age<80.5	7,911	20,961					
					11 Am2	1101	mot>38.5	460	8,909
2 Bdt	0201	mot>51.5 & cog>23.5	355	7,024	11 Am2	1102	mot<38.5	344	13,151
2 Bdt	0202	mot>37.5 & mot<51.5 & cog>32.5	112	7,730					
2 Bdt	0203	mot>37.5 & mot<51.5 & cog>23.5 & cog<32.5	641	9,644	12 Ar1	1201	mot>40.5	4,506	8,482
2 Bdt	0204	mot>46.5 & cog<23.5	505	10,034	12 Ar1	1202	mot<40.5	3,611	11,272
2 Bdt	0205	mot>37.5 & mot<46.5 & cog<23.5	558	12,209					
2 Bdt	0206	mot>25.5 & mot<37.5	953	13,934	13 Ar2	1301	mot>37.5	2,159	8,257
2 Bdt	0207	mot<25.5	653	18,322	13 Ar2	1302	mot>27.5 & mot<37.5	878	10,956
					13 Ar2	1303	mot<27.5	504	13,871
3 Bdn	0301	mot>61.5	160	6,798					
3 Bdn	0302	mot>43.5 & mot<61.5 & cog>22.5	1,373	8,683	14 Car	1401	mot>54.5	2,203	6,338
3 Bdn	0303	mot>43.5 & mot<61.5 & cog<22.5	812	10,295	14 Car	1402	mot>43.5 & mot<54.5	6,456	7,986
3 Bdn	0304	mot>38.5 & mot<43.5	946	11,331	14 Car	1403	mot>35.5 & mot<43.5	5,343	9,589
3 Bdn	0305	mot>28.5 & mot<38.5	1,509	12,941	14 Car	1404	mot<35.5	4,083	12,160
3 Bdn	0306	mot<28.5	1,346	16,969					
					15 Pul	1501	mot>58.5	802	7,370
4 Sct	0401	mot>44.5	307	8,405	15 Pul	1502	mot>44.5 & mot<58.5	3,085	8,746
4 Sct	0402	mot>32.5 & mot<44.5	440	11,277	15 Pul	1503	mot>29.5 & mot<44.5	2,988	11,121
4 Sct	0403	mot>25.5 & mot<32.5	249	15,575	15 Pul	1504	mot<29.5	732	15,031
4 Sct	0404	mot<25.5 & age>78.5	102	19,122					
4 Sct	0405	mot<25.5 & age>64.5 & age<78.5	238	25,881	16 Pai	1601	mot>41.5	4,084	7,139
4 Sct	0406	mot<25.5 & age<64.5	191	18,624	16 Pai	1602	mot>28.5 & mot<41.5	2,860	9,243

Table B.1—Continued

RIC	FRG	Condition	N	Average Cost ($)	RIC	FRG	Condition	N	Average Cost ($)
					16 Pai	1603	mot<28.5	714	11,791
5 Scn	0501	mot>55.5	664	5,933					
5 Scn	0502	mot>44.5 & mot<55.5	3,031	7,313	17 Mt1	1701	mot>50.5	178	6,716
5 Scn	0503	mot>34.5 & mot<44.5	3,955	9,334	17 Mt1	1702	mot>36.5 & mot<50.5	1,231	9,468
5 Scn	0504	mot>25.5 & mot<34.5	2,133	13,173	17 Mt1	1703	mot>28.5 & mot<36.5	930	11,872
5 Scn	0505	mot<25.5	1,453	18,124	17 Mt1	1704	mot<28.5	1,055	14,356
6 Neu	0601	mot>50.5	2,072	7,644	18 Mt2	1801	mot>35.5	242	9,970
6 Neu	0602	mot>39.5 & mot<50.5	5,151	9,414	18 Mt2	1802	mot<35.5	338	17,227
6 Neu	0603	mot>31.5 & mot<39.5	3,418	11,354					
6 Neu	0604	mot<31.5	3,972	13,616	19 Gbr	1901	mot>36.5	189	9,778
					19 Gbr	1902	mot>25.5 & mot<36.5	109	18,123
7 Or1	0701	mot>45.5	4,640	7,421	19 Gbr	1903	mot<25.5	114	26,068
7 Or1	0702	mot>38.5 & mot<45.5	9,831	9,237					
7 Or1	0703	mot>30.5 & mot<38.5	11,923	11,052	20 Mis	2001	mot>51.5	5,251	6,977
7 Or1	0704	mot<30.5	11,214	13,224	20 Mis	2002	mot>41.5 & mot<51.5	13,227	8,558
					20 Mis	2003	mot>31.5 & mot<41.5	11,782	10,614
8 Or2	0801	mot>55.5	5,332	5,106	20 Mis	2004	mot<31.5	7,767	13,336
8 Or2	0802	mot>41.5 & mot<55.5	43,346	6,320					
8 Or2	0803	mot>32.5 & mot<41.5 & age>85.5	1,489	9,157	21 Bur	2101	mot>31.5	104	12,915
8 Or2	0804	mot>32.5 & mot<41.5 & age<85.5	27,112	7,811	21 Bur	2102	mot<31.5	57	26,118
8 Or2	0805	mot>26.5 & mot<32.5	7,533	9,443					
8 Or2	0806	mot<26.5	3,905	11,105					

Table B.2
FRGs Produced by the 2002 Data, Index Variation Alt2

RIC	FRG	Condition	N	Average Cost ($)	RIC	FRG	Condition	N	Average Cost ($)
1 Str	0101	mot>54.25	2,266	7,110	9 Or3	0901	mot>46.75	1,949	6,486
1 Str	0102	mot>46.25 & mot<54.25 & cog>21.5	5,056	8,235	9 Or3	0902	mot>38.25 & mot<46.75	5,460	7,937
1 Str	0103	mot>46.25 & mot<54.25 & cog<21.5	1,714	9,586	9 Or3	0903	mot>29.75 & mot<38.25	4,894	9,598
1 Str	0104	mot>40.75 & mot<46.25	9,126	10,418	9 Or3	0904	mot<29.75	3,352	11,708
1 Str	0105	mot>36.25 & mot<40.75	7,179	12,194					
1 Str	0106	mot>29.25 & mot<36.25	9,686	14,623	10 Am1	1001	mot>45.75	1,481	8,945
1 Str	0107	mot>24.25 & mot<29.25	5,856	17,622	10 Am1	1002	mot>38.25 & mot<45.75	2,160	11,169
1 Str	0108	mot<24.25 & age>80.5	3,443	18,137	10 Am1	1003	mot<38.25	4,382	14,402
1 Str	0109	mot<24.25 & age<80.5	7,967	20,948					
					11 Am2	1101	mot>37.25	458	8,873
2 Bdt	0201	mot>50.75 & cog>23.5	262	6,696	11 Am2	1102	mot<37.25	346	13,191
2 Bdt	0202	mot>36.75 & mot<50.75 & cog>23.5	814	9,107					
2 Bdt	0203	mot>44.25 & cog<23.5	484	9,925	12 Ar1	1201	mot>38.75	4,584	8,502
2 Bdt	0204	mot>36.75 & mot<44.25 & cog<23.5	532	12,155	12 Ar1	1202	mot>34.75 & mot<38.75	1,205	10,156
2 Bdt	0205	mot>27.25 & mot<36.75	818	13,400	12 Ar1	1203	mot<34.75	2,328	11,958
2 Bdt	0206	mot<27.25	867	17,572					
					13 Ar2	1301	mot>36.25	2,120	8,256
3 Bdn	0301	mot>43.75 & cog>23.5	1,104	8,196	13 Ar2	1302	mot>24.75 & mot<36.25	1,027	10,988
3 Bdn	0302	mot>43.75 & cog<23.5	770	9,847	13 Ar2	1303	mot<24.75	394	14,306
3 Bdn	0303	mot>37.25 & mot<43.75	1,351	10,794					
3 Bdn	0304	mot>28.25 & mot<37.25	1,414	12,765	14 Car	1401	mot>52.75	1,851	6,214
3 Bdn	0305	mot>12.5 & mot<28.25	1,391	16,229	14 Car	1402	mot>43.25 & mot<52.75	5,540	7,748
3 Bdn	0306	mot<12.5	116	22,970	14 Car	1403	mot>34.25 & mot<43.25	6,520	9,421
					14 Car	1404	mot<34.25	4,174	12,117
4 Sct	0401	mot>43.25	279	8,180					
4 Sct	0402	mot>30.25 & mot<43.25	493	11,283	15 Pul	1501	mot>53.25	1,125	7,454
4 Sct	0403	mot>17.25 & mot<30.25	483	17,823	15 Pul	1502	mot>42.25 & mot<53.25	2,970	8,957
4 Sct	0404	mot<17.25	272	24,029	15 Pul	1503	mot>28.25 & mot<42.25	2,771	11,211

Table B.2—Continued

RIC	FRG	Condition	N	Average Cost ($)
5 Scn	0501	mot>43.25	3,426	6,966
5 Scn	0502	mot>36.25 & mot<43.25	3,014	8,822
5 Scn	0503	mot>33.25 & mot<36.25	1,128	10,264
5 Scn	0504	mot>24.25 & mot<33.25	2,187	13,068
5 Scn	0505	mot<24.25	1,481	18,189
6 Neu	0601	mot>47.25	2,328	7,690
6 Neu	0602	mot>38.25 & mot<47.25	4,705	9,441
6 Neu	0603	mot>30.25 & mot<38.25	3,529	11,279
6 Neu	0604	mot<30.25	4,051	13,614
7 Or1	0701	mot>44.25	4,187	7,324
7 Or1	0702	mot>37.25 & mot<44.25	10,038	9,159
7 Or1	0703	mot>29.25 & mot<37.25	12,097	11,041
7 Or1	0704	mot<29.25	11,286	13,219
8 Or2	0801	mot>51.25	7,143	5,201
8 Or2	0802	mot>39.25 & mot<51.25	45,081	6,430
8 Or2	0803	mot>30.25 & mot<39.25 & age>83.5	2,597	9,032
8 Or2	0804	mot>30.25 & mot<39.25 & age<83.5	24,092	7,910
8 Or2	0805	mot>24.25 & mot<30.25	6,661	9,686
8 Or2	0806	mot<24.25	3,143	11,275

RIC	FRG	Condition	N	Average Cost ($)
15 Pul	1504	mot<28.25	741	14,962
16 Pai	1601	mot>40.25	3,974	7,116
16 Pai	1602	mot>33.25 & mot<40.25	2,010	8,830
16 Pai	1603	mot<33.25	1,674	10,732
17	1701	mot>48.75	151	6,480
17 Mt1	1702	mot>34.75 & mot<48.75	1,304	9,470
17 Mt1	1703	mot>27.25 & mot<34.75	878	11,956
17 Mt1	1704	mot<27.25	1,061	14,342
18	1801	mot>32.25	268	10,236
18 Mt2	1802	mot>20.75 & mot<32.25	180	15,081
18 Mt2	1803	mot<20.75	132	21,803
19 Gbr	1901	mot>35.25	186	9,699
19 Gbr	1902	mot>20.75 & mot<35.25	148	19,098
19 Gbr	1903	mot<20.75	78	27,790
20 Mis	2001	mot>47.25	7,163	7,151
20 Mis	2002	mot>39.25 & mot<47.25	12,366	8,796
20 Mis	2003	mot>30.25 & mot<39.25	10,620	10,707
20 Mis	2004	mot<30.25	7,878	13,309
21 Bur	2101	mot>29.75	105	12,918
21 Bur	2102	mot<29.75	56	26,435

Table B.3
FRGs Produced by the 2002 Data, Index Variation Alt3

RIC	FRG	Condition	N	Average Cost ($)	RIC	FRG	Condition	N	Average Cost ($)
1 Str	0101	mot>65.75	2,648	7,112	9 Or3	0901	mot>57.75	2,320	6,555
1 Str	0102	mot>57.75 & mot<65.75	6,525	8,684	9 Or3	0902	mot>47.75 & mot<57.75	5,878	8,126
1 Str	0103	mot>52.25 & mot<57.75	7,849	10,292	9 Or3	0903	mot>38.75 & mot<47.75	4,213	9,736
1 Str	0104	mot>46.75 & mot<52.25	7,601	12,061	9 Or3	0904	mot<38.75	3,244	11,794
1 Str	0105	mot>42.25 & mot<46.75	5,633	13,865					
1 Str	0106	mot>36.25 & mot<42.25	6,707	15,965	10 Am1	1001	mot>60.25	1,008	8,504
1 Str	0107	mot<36.25 & age>80.5	4,661	17,676	10 Am1	1002	mot>48.75 & mot<60.25	2,745	10,937
1 Str	0108	mot>31.25 & mot<36.25 & age<80.5	3,374	18,550	10 Am1	1003	mot<48.75	4,270	14,510
1 Str	0109	mot<31.25 & age<80.5	7,295	20,863					
					11 Am2	1101	mot>47.75	458	8,930
2 Bdt	0201	mot>59.25 & cog>31.5	129	5,917	11 Am2	1102	mot<47.75	346	13,081
2 Bdt	0202	mot>59.25 & cog>23.5 & cog<31.5	305	7,846					
2 Bdt	0203	mot>45.75 & mot<59.25 & cog>23.5	688	9,480	12 Ar1	1201	mot>62.25	965	7,563
2 Bdt	0204	mot>58.75 & cog<23.5	287	9,492	12 Ar1	1202	mot>46.75 & mot<62.25	4,310	8,947
2 Bdt	0205	mot>45.75 & mot<58.75 & cog<23.5	728	11,711	12 Ar1	1203	mot<46.75	2,842	11,674
2 Bdt	0206	mot>32.75 & mot<45.75	912	13,762					
2 Bdt	0207	mot<32.75	728	18,160	13 Ar2	1301	mot>56.75	854	7,353
					13 Ar2	1302	mot>45.75 & mot<56.75	1,338	9,042
3 Bdn	0301	mot>54.25 & cog>22.5	1,241	8,308	13 Ar2	1303	mot<45.75	1,349	11,849
3 Bdn	0302	mot>54.25 & cog<22.5	657	9,879					
3 Bdn	0303	mot>45.75 & mot<54.25	1,447	10,921	14 Car	1401	mot>59.25	4,305	6,800
3 Bdn	0304	mot>31.75 & mot<45.75	1,766	13,501	14 Car	1402	mot>49.25 & mot<59.25	6,804	8,590
3 Bdn	0305	mot>20.25 & mot<31.75	779	16,271	14 Car	1403	mot>43.25 & mot<49.25	3,047	10,003
3 Bdn	0306	mot<20.25	256	20,860	14 Car	1404	mot<43.25	3,929	12,293
4 Sct	0401	mot>53.25	312	8,152	15 Pul	1501	mot>55.25	3,567	8,343
4 Sct	0402	mot>40.25 & mot<53.25	421	11,311	15 Pul	1502	mot>37.25 & mot<55.25	3,240	10,923
4 Sct	0403	mot>30.75 & mot<40.25	312	15,952	15 Pul	1503	mot<37.25	800	14,881
4 Sct	0404	mot<30.75 & age>57.5	331	24,438					
4 Sct	0405	mot<30.75 & age<57.5	151	18,323	16 Pai	1601	mot>50.25	4,248	7,180
					16 Pai	1602	mot>42.75 & mot<50.25	1,840	8,957

Table B.3—Continued

RIC	FRG	Condition	N	Average Cost ($)	RIC	FRG	Condition	N	Average Cost ($)
5 Scn	0501	mot>62.75	954	6,082	16 Pai	1603	mot<42.75	1,570	10,833
5 Scn	0502	mot>52.75 & mot<62.75	3,161	7,483					
5 Scn	0503	mot>45.25 & mot<52.75	2,704	9,143	17 Mt1	1701	mot>61.75	126	6,373
5 Scn	0504	mot>42.25 & mot<45.25	886	10,582	17 Mt1	1702	mot>54.75 & mot<61.75	378	8,223
5 Scn	0505	mot>39.25 & mot<42.25	749	11,633	17 Mt1	1703	mot>46.25 & mot<54.75	850	9,858
5 Scn	0506	mot>32.75 & mot<39.25	1,259	14,304	17 Mt1	1704	mot>38.25 & mot<46.25	840	11,604
5 Scn	0507	mot<32.75	1,523	18,092	17 Mt1	1705	mot<38.25	1,200	14,199
6 Neu	0601	mot>57.25	2,943	7,839	18 Mt2	1801	mot>53.25	101	8,565
6 Neu	0602	mot>47.25 & mot<57.25	4,566	9,710	18 Mt2	1802	mot>37.75 & mot<53.25	230	11,929
6 Neu	0603	mot>39.75 & mot<47.25	2,844	11,268	18 Mt2	1803	mot<37.75	249	18,876
6 Neu	0604	mot<39.75	4,260	13,580					
					19 Gbr	1901	mot>45.25	191	9,674
7 Or1	0701	mot>54.75	5,394	7,519	19 Gbr	1902	mot<45.25	221	22,188
7 Or1	0702	mot>47.25 & mot<54.75	9,687	9,355					
7 Or1	0703	mot>39.25 & mot<47.25	10,585	11,059	20 Mis	2001	mot>59.25	6,647	7,096
7 Or1	0704	mot<39.25	11,942	13,161	20 Mis	2002	mot>48.25 & mot<59.25	14,249	8,878
					20 Mis	2003	mot>36.75 & mot<48.25	10,614	10,966
8 Or2	0801	mot>62.75	8,235	5,258	20 Mis	2004	mot<36.75	6,517	13,559
8 Or2	0802	mot>49.75 & mot<62.75 & age>85.5	1,585	7,504					
8 Or2	0803	mot>49.75 & mot<62.75 & age<85.5	44,411	6,459	21 Bur	2101	mot>41.75	96	12,608
8 Or2	0804	mot>39.25 & mot<49.75	25,812	8,091	21 Bur	2102	mot<41.75	65	24,816
8 Or2	0805	mot>31.75 & mot<39.25	6,267	9,886					
8 Or2	0806	mot<31.75	2,407	11,689					

Table B.4
FRGs Produced by the 2002 Data, Index Variation Alt4

RIC	FRG	Condition	N	Average Cost ($)	RIC	FRG	Condition	N	Average Cost ($)
1 Str	0101	mot>53.35	3,135	7,219	9 Or3	0901	mot>47.05	2,087	6,488
1 Str	0102	mot>46.35 & mot<53.35	6,638	8,860	9 Or3	0902	mot>36.65 & mot<47.05	6,477	8,120
1 Str	0103	mot>40.75 & mot<46.35	9,097	10,488	9 Or3	0903	mot>28.95 & mot<36.65	3,993	9,863
1 Str	0104	mot>37.15 & mot<40.75	5,590	12,237	9 Or3	0904	mot<28.95	3,098	11,864
1 Str	0105	mot>32.55 & mot<37.15	6,521	13,930					
1 Str	0106	mot>28.45 & mot<32.55	5,242	15,781	10 Am1	1001	mot>45.35	1,640	9,025
1 Str	0107	mot<28.45 & age>80.5	4,844	17,744	10 Am1	1002	mot>37.45 & mot<45.35	2,255	11,344
1 Str	0108	mot>23.05 & mot<28.45 & age<80.5	4,262	18,641	10 Am1	1003	mot<37.45	4,128	14,578
1 Str	0109	mot<23.05 & age<80.5	6,964	21,065					
					11 Am2	1101	mot>36.9	457	8,840
2 Bdt	0201	mot>49.45 & cog>31.5	109	5,765	11 Am2	1102	mot<36.9	347	13,242
2 Bdt	0202	mot>49.45 & cog>23.5 & cog<31.5	252	7,669					
2 Bdt	0203	mot>49.45 & cog<23.5	221	9,340	12 Ar1	1201	mot>51.05	909	7,519
2 Bdt	0204	mot>35.15 & mot<49.45 & cog>31.5	158	8,190	12 Ar1	1202	mot>38.15 & mot<51.05	3,866	8,815
2 Bdt	0205	mot>35.15 & mot<49.45 & cog>19.5 & cog<31.5	965	9,974	12 Ar1	1203	mot<38.15	3,342	11,399
2 Bdt	0206	mot>35.15 & mot<49.45 & cog<19.5	548	12,536					
2 Bdt	0207	mot>24.05 & mot<35.15	873	14,094	13 Ar2	1301	mot>36.65	2,077	8,217
2 Bdt	0208	mot<24.05	651	18,482	13 Ar2	1302	mot>25.65 & mot<36.65	993	10,827
					13 Ar2	1303	mot<25.65	471	14,061
3 Bdn	0301	mot>43.55 & cog>22.5	1,258	8,330					
3 Bdn	0302	mot>43.55 & cog<22.5	724	9,929	14 Car	1401	mot>51.95	2,337	6,370
3 Bdn	0303	mot>37.35 & mot<43.55	1,227	10,878	14 Car	1402	mot>42.65 & mot<51.95	5,859	7,911
3 Bdn	0304	mot>26.75 & mot<37.35	1,615	12,932	14 Car	1403	mot>33.85 & mot<42.65	5,938	9,566
3 Bdn	0305	mot>12.65 & mot<26.75	1,212	16,555	14 Car	1404	mot<33.85	3,951	12,270
3 Bdn	0306	mot<12.65	110	23,180					
					15 Pul	1501	mot>58.05	587	7,191
4 Sct	0401	mot>44.05	265	8,050	15 Pul	1502	mot>43.65 & mot<58.05	3,150	8,610
4 Sct	0402	mot>31.45 & mot<44.05	459	10,978	15 Pul	1503	mot>30.55 & mot<43.65	2,884	10,962
4 Sct	0403	mot>22.15 & mot<31.45	344	16,106	15 Pul	1504	mot<30.55	986	14,374
4 Sct	0404	mot<22.15 & age>64.5	283	25,079					
4 Sct	0405	mot<22.15 & age<64.5	176	18,845	16 Pai	1601	mot>40.75	3,864	7,081

Table B.4—Continued

RIC	FRG	Condition	N	Average Cost ($)	RIC	FRG	Condition	N	Average Cost ($)
					16 Pai	1602	mot>32.65 & mot<40.75	2,266	8,827
5 Scn	0501	mot>53.35	622	5,868	16 Pai	1603	mot<32.65	1,528	10,905
5 Scn	0502	mot>43.55 & mot<53.35	2,802	7,216					
5 Scn	0503	mot>36.65 & mot<43.55	2,878	8,760	17 Mt1	1701	mot>44.95	392	7,469
5 Scn	0504	mot>32.65 & mot<36.65	1,426	10,317	17 Mt1	1702	mot>35.25 & mot<44.95	1,017	9,693
5 Scn	0505	mot>24.75 & mot<32.65	1,858	13,118	17 Mt1	1703	mot>27.05 & mot<35.25	949	11,886
5 Scn	0506	mot<24.75	1,650	17,913	17 Mt1	1704	mot<27.05	1,036	14,500
6 Neu	0601	mot>45.65	3,196	7,916	18 Mt2	1801	mot>35.25	229	9,799
6 Neu	0602	mot>35.65 & mot<45.65	5,096	9,879	18 Mt2	1802	mot>20.95 & mot<35.25	211	14,496
6 Neu	0603	mot>29.95 & mot<35.65	2,341	11,618	18 Mt2	1803	mot<20.95	140	21,849
6 Neu	0604	mot<29.95	3,980	13,690					
					19 Gbr	1901	mot>34.5	193	9,795
7 Or1	0701	mot>42.65	5,876	7,568	19 Gbr	1902	mot>23.15 & mot<34.5	109	18,555
7 Or1	0702	mot>36.65 & mot<42.65	8,702	9,372	19 Gbr	1903	mot<23.15	110	26,310
7 Or1	0703	mot>28.85 & mot<36.65	11,591	11,033					
7 Or1	0704	mot<28.85	11,439	13,256	20 Mis	2001	mot>47.35	7,461	7,171
					20 Mis	2002	mot>39.55 & mot<47.35	11,692	8,790
8 Or2	0801	mot>50.95	8,161	5,232	20 Mis	2003	mot>28.65 & mot<39.55	12,134	10,790
8 Or2	0802	mot>38.75 & mot<50.95 & age>85.5	1,568	7,467	20 Mis	2004	mot<28.65	6,740	13,564
8 Or2	0803	mot>38.75 & mot<50.95 & age<85.5	43,376	6,433					
8 Or2	0804	mot>30.25 & mot<38.75	25,009	8,005	21 Bur	2101	mot>29.95	103	12,956
8 Or2	0805	mot>24.25 & mot<30.25	7,052	9,592	21 Bur	2102	mot<29.95	58	25,658
8 Or2	0806	mot<24.25	3,551	11,223					

Bibliography

Breiman, L., J. Friedman, R. Olshen, and C. Stone, *Classification and Regression Trees*, Belmont, Calif.: Wadsworth, 1984.

Carter, Grace M., and Mark Totten, *Preliminary Analyses for Refinement of the Tier Comorbidities in the IRF PPS,* Santa Monica, Calif.: RAND Corporation, TR-201-CMS, forthcoming.

Hastie, T., R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning,* New York, N.Y.: Springer-Verlag, 2001.

Newhouse, J., S. Cretin, and C. Witsberger, "Predicting Hospital Accounting Costs," *Health Care Financing Review,* Vol. 11, No. 1, pp. 25–33.

Stineman M., B. Hamilton, C. Granger, J. Goin, J. Escarce, and S. Williams, "Four Methods for Characterizing Disability in the Formation of FRGs," *Archives of Physical Medicine and Rehabilitation,* Vol. 75, December 1994, pp. 1277–1283.