RAND

Final Report on Assessment Instruments for a Prospective Payment System

Joan L. Buchanan, Patricia Andres, Stephen M. Haley, Susan M. Paddock, David C. Young, Alan Zaslavsky

Prepared for the
Centers for Medicare and Medicaid Services

RAND Health

The research described in this report was sponsored by the Centers for Medicare and Medicaid Services (formerly the Health Care Financing Administration). The research was conducted through a subcontract from RAND to Harvard University and represents a collaborative effort involving faculty from the department of Health Care Policy at Harvard Medical School, Sargent College of Health and Rehabilitation Sciences at Boston University and RAND Health.

Library of Congress Cataloging-in-Publication Data

Final report on assessment instruments for prospective payment system / Joan L. Buchanan ...
[et al.].
p. cm.
"MR-1501."
Includes bibliographical references.
ISBN 0-8330-3221-6
1. Hospitals—Rehabilitation services—Prospective payment. I. Buchanan, Joan, 1947–

RA971.32 .F564 2002
362.1'786—dc21

2002031821

The RAND Corporation is a nonprofit research organization providing objective analysis and effective solutions that address the challenges facing the public and private sectors around the world. RAND's publications do not necessarily reflect the opinions of its research clients and sponsors.

RAND® is a registered trademark.

A profile of RAND Health, abstracts of its publications, and ordering information can be found on the RAND Health home page at www.rand.org/health.

© Copyright 2004 RAND Corporation

All rights reserved. No part of this book may be reproduced in any form by any electronic or mechanical means (including photocopying, recording, or information storage and retrieval) without permission in writing from RAND.

Published 2004 by the RAND Corporation
1700 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

The Balanced Budget Act of 1997 mandated the implementation of a prospective payment system for inpatient rehabilitation. The Health Care Financing Administration (now the Centers for Medicare and Medicaid Services) issued a Notice of Proposed Rule Making in the *Federal Register* on November 3, 2000, which described the design of the rehabilitation prospective payment system (PPS) and reflected their desire to substitute a new, broader, multipurpose data collection instrument, the Minimum Data Set—Post-Acute Care, for the original Functional Independence Measure. This study assesses the potential implications of making this substitution on patient classification and facility payment.

The appendices to this report are published in a separate volume as J. L. Buchanan, P. Andres, S. M. Haley, S. M. Paddock, D. C. Young, and A. Zaslavsky, *Final Report on Assessment Instruments for a Prospective Payment System: Appendices,* Santa Monica, CA: RAND, MR-1501/1-CMS, 2002.

This report is part of a series of RAND reports describing the analytic work that underlies the design of the rehabilitation hospital PPS. Other reports in that series include:

G. M. Carter, J. L. Buchanan, M. B. Buntin, O. Hayden, S. M. Paddock, J. H. Kawata, D. A. Relles, G. K. Ridgeway, M. Totten, and B. O. Wynn, *Executive Summary of Analyses for the Initial Implementation of the Inpatient Rehabilitation Facility Prospective Payment System*, Santa Monica, CA: RAND, MR-1500/1-CMS, 2002.

G. M. Carter, M. B. Buntin, O. Hayden, J. H. Kawata, S. M. Paddock, D. A. Relles, G. K. Ridgeway, M. Totten, and B. O. Wynn, *Analyses for the Initial Implementation of the Inpatient Rehabilitation Facility Prospective Payment System,* Santa Monica, CA: RAND, MR-1500-CMS, 2002.

D. A. Relles and G. M. Carter, *Linking Medicare and Rehabilitation Hospital Records to Support Development of a Rehabilitation Hospital Prospective Payment System,* Santa Monica, CA: RAND, MR-1502-CMS.

G. M. Carter, D. A. Relles, B. O. Wynn, J. Kawata, S. M. Paddock, N. Soon, and M. E. Totten, *Interim Report on an Inpatient Rehabilitation Facility Prospective Payment System,* Santa Monica, CA: RAND, MR-1503-CMS.

The research described here was a collaborative effort undertaken by the Department of Health Care Policy, Harvard Medical School, Sargent College of Health and Rehabilitation Sciences at Boston University, and others to support RAND and the Centers for Medicare and Medicaid Services in their efforts to design and implement a prospective payment system for inpatient rehabilitation. The role of each organization is described below.

Harvard Medical School had responsibility for the project design, its implementation, analysis, and the preparation of the final report.

Sargent College of Health and Rehabilitation Sciences at Boston University provided the rehabilitation expertise for the development and preparation of training materials, hired the calibration teams, supervised their training, and conducted the certification process for both the calibration teams and the institutional-based data collection teams. They were responsible for hiring the field core coordinator who oversaw the entire field operation. They also participated in the analytic phase of the project.

RAND was the prime contractor providing both overall guidance and some of the analytic support to the project, particularly for the factor analysis.

Hebrew Rehabilitation Center for the Aged conducted the train the trainers program on the MDS-PAC and ran a regular MDS-PAC training program for our calibration teams. They also provided expert consultation to the field coordinator and the study team on questions with the MDS-PAC.

Uniform Data System for Medical Rehabilitation provided trainers and conducted 10 training sessions across the country on the MDS-PAC for our institutional data collectors. They also conducted the FIM training session for the study's calibration teams. They provided expert consultation to the field coordinator and the study team on FIM scoring rules.

Contents

Preface iii

Figures vii

Tables ix

Summary xi

Acknowledgments xxi

Acronyms xxiii

1. INTRODUCTION 1
   - Background on the Development of a Prospective Payment System for Inpatient Rehabilitation 1
   - Background on Instrument Performance 3
   - Purpose and Scope of This Project 5
   - Organization of the Report 6

2. STUDY DESIGN AND IMPLEMENTATION 7
   - Research Questions 7
   - Sample Size 7
   - Hospital Recruitment 8
   - Facility Selection 9
   - Training the Trainers 10
   - Facility Training 12
   - Data Collection and Transmission 13
   - Monitoring and Communication with the Field 13
   - Calibration Teams 14

3. TRANSLATING THE MDS-PAC INTO FIM MOTOR AND COGNITIVE SCALE ITEMS 16
   - The Morris Translation 18
   - Admission Translation Rationale 21
     - Use of ADL Assist Codes 22
     - Scoring Independence 23
     - Scoring Setup and Supervision 23
   - Item-by-Item Translation 23
     - Grooming, Bathing, Dressing Upper Body, Dressing Lower Body, Toileting, Transfer Toilet , and Transfer Tub/Shower 24
     - Eating 26
     - Bladder Management 29
     - Bowel Management 34
     - Transfer Bed, Chair, Wheelchair 36
     - Locomotion 36
     - Stairs 41
   - Scoring Differences That Could Not Be Corrected in the Translation 41

General Limitations 41
Limitations in Converting MDS-PAC to FIM Scores for Specific Items 43
Additional Comments 44
Summary 45

4. CONCLUSIONS 46
Coding the Reason for a Rehabilitation Admission 46
Completeness of Other Items 50
Rescoring Reliabilities Between Institutional and Calibration Teams on the FIM and the MDS-PAC 51
Are the FIM and the MDS-PAC Measuring the Same Concepts? 57
How Items Within the MDS-PAC Cluster to Form Common Factors 58
How the Combined FIM and the MDS-PAC Raw Items Load onto Common or Distinct Factors 61
Comparing Translations of MDS-PAC-Based Items with FIM 66
Instrument Completion Times 71
Summary 73

5. ACCURACY OF THE MDS-PAC TRANSLATION INTO PSEUDO-FIM ITEMS 75
Introduction 75
What Factors Contribute to These Observed Differences? 80
Differences in the Assessment Periods 80
Other Possible Contributors 81
The Role of Scoring Error 84
Summary 85

6. MAPPING PSEUDO-FIM MOTOR AND COGNITIVE SCORES INTO CMGS 87
Introduction 87
CMGs 87
Mapping and Adjusting Pseudo-FIM Scores to Match FIM Scores 88
Accuracy of Alternative Mappings 90
Payment Differences 92
Regression Analysis of Payment Differences 96
Summary 98

REFERENCES 101

Figures

2.1. Map of Selected Facilities 10
3.1. FIM Scoring 17
3.2. MDS-PAC Section E: Functional Status 19
3.3. MDS-PAC Section F: Bladder/Bowel Management 20

Tables

2.1. Sample Size . . . 8
2.2. Characteristics of Selected Facilities . . . 10
3.1. Comparison of Scale and Item Means Using the Morris Translation . . . 21
3.2. Grooming . . . 25
3.3. Bathing . . . 26
3.4. Dressing Upper Body . . . 27
3.5. Dressing Lower Body . . . 28
3.6. Toileting . . . 29
3.7. Transfer Toilet . . . 30
3.8. Transfer Tub/Shower . . . 31
3.9. Eating . . . 32
3.10. Bladder Management . . . 33
3.11. Bowel Management . . . 35
3.12. Transfer Bed, Chair, Wheelchair . . . 37
3.13. Locomotion—Walk/Wheelchair . . . 38
3.14. Stairs . . . 41
4.1. Number and Percentage of Cases with Usable Impairment Code Data . . . 48
4.2. Disagreement on RIC Selection by Instrument and Scoring Team Combination . . . 49
4.3. RIC Agreement Between Institutional and Calibration Teams . . . 50
4.4. Number and Percentage of Cases with Missing Functional and Cognitive Status Data . . . 51
4.5. Rescoring Reliabilities Between Institutional and Calibration Teams: Pearson Correlation Coefficients . . . 53
4.6. Rescoring Reliabilities Between Institutional and Calibration Teams: Kappa Statistics . . . 54
4.7. Comparisons of Absolute Agreement Between Institutional and Calibration Teams on the FIM and the MDS-PAC . . . 56
4.8. Eigenvalues of Factors for the Unrotated Factor Analyses of Each MDS-PAC Section . . . 58
4.9. Factors in Sections B–F of the MDS-PAC . . . 60
4.10. Top 20 Eigenvalues for Unrotated Factor Analysis of Combined FIM and Raw MDS-PAC Cognitive Items . . . 62
4.11. Factors Resulting When MDS-PAC and FIM Cognitive Items Are Combined . . . 63
4.12. Communalities on Raw MDS-PAC and FIM Functional Status Items . . . 63
4.13. Top 20 Eigenvalues for Unrotated Factor Analysis of Combined FIM and Raw MDS-PAC Motor Scale Items . . . 64
4.14. Thirteen Factors Identified in the Analysis of Combined FIM and MDS-PAC Motor Scale Items . . . 65
4.15. Top 10 Eigenvalues for Unrotated Factors in the Analysis of Pseudo-FIM and FIM Item Scores . . . 66

4.16. Eight Factors Identified in the Analysis of Pseudo-FIM and FIM Scores 67
4.17. Cluster Analysis Output for Combined Motor and Cognitive Items 67
4.18. Cluster Definitions 68
4.19. Top 10 (of 26) Eigenvalues for Unrotated Factors of Pseudo-FIM and FIM Motor Items 70
4.20. Top 10 Eigenvalues for Unrotated Factors of the Morris Translation and FIM Motor Items 70
4.21. Six Factors Derived from the Factor Analysis of Pseudo-FIM and FIM Motor Scores 70
4.22. Five Factors Derived from the Factor Analysis of the Morris Translation and FIM Motor Scores 70
4.23. Average Time Required to Complete Each Instrument 72
4.24. MDS-PAC Completion Times by Period and Team Size 73
4.25. Regression Model Explaining Log (MDS-PAC Completion Time) 74
5.1. Comparison of Motor and Cognitive Scales and Item-Level Means Across Instruments and Samples 75
5.2. Comparison of Motor and Cognitive Scale Distributions Across Instruments and Samples 76
5.3. Comparison of Pearson Correlation Coefficients Across Instruments and Samples 77
5.4. Comparison of Motor and Cognitive Scales and Item-Level Kappas Across Instruments and Samples 78
5.5. Comparison of Motor and Cognitive Scales and Item-Level Agreement Across Instruments and Samples 79
5.6. Comparison of Motor and Cognitive Scale Means and Distributions Across Instruments by FIM Assessment Day 81
5.7. MDS-PAC and FIM Agreement by FIM Assessment Day 82
5.8. Regression Models for Motor Score Differences (PAC-FIM) 84
6.1. Adjustments to Standardize Pseudo-FIM Motor and Cognitive Scales to Actual FIM Motor and Cognitive Scale Means and Standard Deviations 90
6.2. Regression Coefficients for Transforming Pseudo-FIM Motor and Cognitive Items to Actual FIM Motor and Cognitive Scales 91
6.3. Agreement and Disagreement on Case Mix Group Classification Within RICs Between the FIM and the MDS-PAC 91
6.4. Agreement at the RIC and CMG Level 93
6.5. Within-Instrument Agreement and Disagreement on Case Mix Group Classification Within RICs 96
6.6. Means and Standard Deviations of Payment Levels 97
6.7. Regression Models on Payment Differences and Absolute Payment Differences 99

Summary

The Balanced Budget Act of 1997 mandated the implementation of three prospective payment systems for post-acute care providers—one for nursing homes, another for home health agencies, and a third for inpatient rehabilitation facilities. Prospective payment systems pay providers a predetermined fixed price (per day, per episode, or per case) that depends on patient resource needs (often a disease profile or reason for admission) but is independent of the amount of services actually provided. Since the payment is independent of service provision, such systems are thought to create an incentive for efficient, cost conscious care. Although the populations being treated in each post-acute setting have many similarities, the new payment systems have little in common. Each is based on different case mix measures from different assessment tools and, further, each uses different levels of aggregation for payment. The new rehabilitation PPS uses the rehabilitation impairment category (a broad grouping of those admitted for similar rehabilitation reasons), patient age, and functional and cognitive status to classify patients and a single payment is made for the admission. The initial design work for this PPS was based on a functional assessment tool, called the Functional Independence Measure (FIM) and a patient classification system called the Functional Independence Measure-Function Related Groups (FIM-FRGs). RAND researchers refined, completed, and updated that classification work and designed the payment system (see Carter et al., 2002a, 2002b, 2002c). As time passed, policymakers increasingly realized their need for cross setting comparisons of the populations being cared for, the treatments being given, and the outcomes. A new assessment tool, similar to that used in the nursing home industry, the Minimum Data Set—Post-Acute Care (MDS-PAC), was developed to replace the FIM in the rehabilitation PPS. This study was undertaken to evaluate the implications of that substitution.

The MDS-PAC is a comprehensive data collection tool with over 300 items including sections on sociodemographic information, pre-admission history, advance directives, cognitive patterns, communication patterns, mood and behavior patterns, functional status, bladder/bowel management, diagnoses, medical complexities, pain status, oral/nutritional status, procedures/services, functional prognosis, and resources for discharge. Data collectors are instructed to interview the patient and family members and to talk to all caregivers over all shifts for the first 72 hours of care as well as to consult the patient's chart. Functional status assessments allow for one or two exceptions where more care is

needed. The MDS-PAC explicitly recognizes that an activity may not have occurred.

In contrast, the typical FIM form contains a short list of items asking for sociodemographic information, an item asking for the impairment group (reason for the rehabilitation admission) and its underlying etiologic diagnosis, and 18 FIM motor and cognitive items scored at both admission and discharge. The instrument must be scored sometime in the first 72 hours after admission (and within 72 hours before discharge) but is generally scored for the most recent 24-hour period. Scoring on the 18 FIM items is usually evaluated by therapists within their areas of expertise. All items must be scored. Any patient who cannot safely perform an activity is automatically scored as totally dependent.

The planned payment system organizes patients into rehabilitation impairment categories based on the therapeutic reason for admission and then uses the FIM motor scale (sum of the 13 motor item scores), the FIM cognitive scale (sum of the five cognitive item scores), and patient age to classify cases into case mix groups (CMGs) for payment. The age, motor, and cognitive scale values that define each payment cell within a rehabilitation impairment category were defined using classification and regression tree analysis. The CMGs used in this report are available in the Notice of Proposed Rule Making, *Federal Register*, November 3, 2000. These have been further refined and the definitions for the final CMGs can be found in Carter et al. (2002a, 2002b).

To use the MDS-PAC in the new payment system, we needed a method to create a FIM-like motor score and a FIM-like cognitive score. Since the basic FIM concepts were embodied in both instruments, we began with a translation that took several items from the MDS-PAC and converted them into 18 FIM-like items. By summing the 13 "pseudo-FIM" motor items from the MDS-PAC, a motor scale was created. Similarly, the five pseudo-FIM cognitive items were created and summed to form a cognitive scale.

The goal of this project was to compare two instruments, the MDS-PAC and the FIM, to provide insight into whether the planned substitution of the MDS-PAC for the FIM in the proposed inpatient rehabilitation hospital prospective payment system would adversely affect system performance, patients, or hospitals.

Study Design and Implementation

The study design called for two types of data collection: (1) institutionally based teams of rehabilitation therapists and nurses collected FIM and MDS-PAC data on all Medicare admissions within a 10-week study time frame, and (2) study-employed data collection teams, also nurses and rehabilitation therapists, traveled to each hospital during the 10-week data collection phase to re-score FIM and MDS-PAC data on a subset of patients. The latter were referred to as calibration teams. The data provided by the institutionally based teams were used for our primary analyses that examined how well the translation of the MDS-PAC into FIM-like items worked and the payment comparisons. The data collected by the calibration teams were used to examine scoring reliability and to see if institutions were scoring to the same set of norms.

All FIM-certified institutions were invited to participate in the study. Potential participants were asked to send one or more teams to a two-day training session to learn how to score the MDS-PAC and were told that training costs would be paid by the study. Institutions were told that they would receive $35 per completed case (MDS-PAC and FIM) up to $4,000. Within a week, the study received over 180 volunteer responses. To facilitate training and limit calibration team travel, all responding facilities were mapped and hospitals in geographic clusters were linked to together. We then created an expected caseload for each cluster using data on the number of Medicare admissions reported during the previous month for each facility in the cluster. This process allowed us to select clusters that geographically spanned the country and had adequate caseload. Consequently, we were able to manage the travel and workload scheduling for the calibration teams and to manage the training of institutionally based data collectors. Six broad regions were selected with 53 hospitals. Three of the selected hospitals could not meet our schedule and were dropped from the study.

FIM and MDS-PAC data were collected on over 3,200 Medicare cases on hand-written forms from the 50 participating rehabilitation units and hospitals. The facilities ranged in size from 13 to 150 beds. Sixteen percent of rehabilitation hospitals were rural and 28 percent were freestanding facilities. Data collectors were teams of clinicians (physical therapists, occupational therapists, speech language pathologists, and nurses) from each site who attended a two-day MDS-PAC training session and successfully completed a certification exam before the start of the study.

Three calibration teams re-rated over 200 of these cases using both the MDS-PAC and the FIM giving us estimates of inter-team scoring reliability. The calibration teams each included a nurse and two therapists at the beginning of the study. Two nurses were lost to the study early in the data collection phase. Before beginning data collection, the calibration teams were formally trained and certified on both the FIM and the MDS-PAC. Then they spent three weeks working intensively together in four rehabilitation hospitals in the greater Boston area. During the 10-week data collection phase, one or more calibration teams visited all study hospitals re-scoring three to eight cases in each hospital.

Study Findings

Translating the MDS-PAC into FIM-Like Items

To classify patients into case mix groups for payment using the MDS-PAC, we needed to create motor and cognitive scales similar to those in the FIM. The FIM motor scale includes 13 items that cover self-care (eating, bathing, grooming, dressing, and toileting), mobility (transfers, locomotion, and stairs), and sphincter control. The FIM cognitive scale has five items (comprehension, expression, social interaction, problem solving, and memory). Each item in these scales is scored from 1 = total assistance to 7 = complete independence.

Like the FIM, the MDS-PAC also includes functional status items covering self-care, mobility, and sphincter control. In the MDS-PAC, these are scored in reverse order with 0 = complete independence and 6 = total assistance. The MDS-PAC uses two questions for each item; one to cover patient self-performance and the other to indicate the level of assistance provided by others. In the FIM, these concepts are combined into a single rating. The MDS-PAC does not have items with obvious parallels to the FIM cognitive items. For the FIM cognitive scale, we used an empirically derived translations of MDS-PAC items into the pseudo-FIM cognitive items that were developed by Dr. John Morris. For the FIM motor scale, we revised his proposed translation of items.

The revised motor scale translation (1) re-aligned the response category mappings often by incorporating information from other parts of the MDS-PAC, (2) incorporated physical assistance more completely into the scoring, and (3) substituted items where this improved performance. Specifically, the revised translation tried to distinguish the concept of modified independence from total independence (the top two categories in the FIM scoring), collapsed setup and supervision into the next level, incorporated the physical assistance items, and tried to correct several other item-specific scoring inconsistencies. The revised translation also substituted the “walk in facility” for the “locomotion” item, since FIM instructions indicate that the locomotion item should be scored for current capability but uses the mode of locomotion expected at discharge and over 85 percent of cases walk at discharge.

Although relatively short, the FIM actually has a fairly complex set of scoring rules, some of which differed explicitly from those in the PAC, and others merely could not be replicated. Among the more obvious differences are (1) the difference in the assessment periods—the MDS-PAC looks back at the first three days after admission and the FIM looks back over 24 hours any time during the first three days; (2) for patients who appear to be independent, the absence of information on the MDS-PAC about whether the task is completed safely and in a reasonable amount of time; (3) the absence of information in the MDS-PAC on one person assistance with the torso or multiple limbs; (4) different definitions of the need for total assistance; and (5) differences in the task definitions and the treatment of medication use for bowel and bladder management.

Evaluating the Translation

We used factor analysis to assess whether the revised translation improved the conceptual agreement between the pseudo-FIM and FIM concepts and found that, in fact, it did. Neither the raw items nor those from the original translation loaded onto the same factors as the corresponding FIM items, but items from the revised translation did.

The revised translation reduced the mean difference in motor scores between the FIM and the MDS-PAC by 50 percent from the original Morris translation. Despite the improvement, we found that the agreement between the instruments for institutionally based scoring teams (as measured by weighted kappa statistics) was only moderate. Absolute agreement (as assessed by simple kappas) was worse, ranging from poor to moderate. However, when the calibration teams scored patients using both instruments, we found notably higher levels of agreement.

We anticipated that differences in the assessment periods between the instruments contributed to the mean difference in motor scores and found, in fact, that they did. Patients whose motor exams were completed on days 1 and 2 had significantly larger differences than those completed on day 3, with day 2 showing the largest difference. Other factors that influenced the difference were the size of the team scoring the MDS-PAC (three-person teams had smaller differences than one-person teams and those with four or more persons after controlling for other variables) and whether the patient was in for lower extremity joint replacement (RIC 8). After controlling explicitly for the variables that we could, we found that a random effect for hospitals was highly significant. The latter implies that hospitals were systematic in their scoring differences and this was not explained by any of the independent variables. This suggests that more training is needed to adequately standardize the assessment process.

Scoring Reliability

Some of the translation difficulties could be attributable to poor scoring reliability within one or both instruments. A well-designed instrument should yield the same or nearly the same scores for a given patient when administered by different teams or individuals. To assess the reliability of the FIM and the PAC, we compared data re-scored by the calibration teams with that collected by the institutional teams.

When we looked at the impairment group item that was the same on both instruments, we found high levels of disagreement between the institutional teams and the calibration teams. We did not compare the impairment groups directly, but rather we employed a weaker test, comparing the RICs that they mapped into and found that 27–29 percent of the time they were invalid or mapped into different RICs. This finding indicated that additional rules or instructions governing RIC selection were needed for both instruments.

When we compared the scoring reliabilities on the FIM and pseudo-FIM items from the FIM and the MDS-PAC, we found that for the motor items, the FIM had modestly higher kappas and levels of absolute agreement than the PAC. However, regardless of which instrument was used, scoring reliabilities on the weighted kappas were generally only moderate (simple kappas showed poor agreement on 8 out of 18 FIM items and 14 out of 18 MDS-PAC items), a concern for measures intended for use in a payment system. Further, our reliability measures for the FIM motor scale, the cognitive scale, and 11 of 13 motor items were less than those reported in a meta analysis of 11 studies in the literature (see Ottenbacher et al., 1996). The inter-team scoring reliabilities in this study fell below the mean, median, and lower confidence limits on the means that they reported for the motor scale, the cognitive scale, and 11 of the 13 motor items. For three of the five cognitive items, our inter-team scoring reliabilities fell between the reported means and medians. For two of the 13 motor items and two of the five cognitive items, our inter-team reliabilities exceeded those reported in the meta analysis. The meta analysis does not provide information on how actual FIM assessments were performed in the 11 studies. Our calibration teams were observers and information gatherers who did not actually do any physical assessment. At times, they were trying to gather information that was as much as three days old. These procedural differences may have contributed to lower scoring reliabilities. However, one could also argue that their greater dependence on information from treating clinicians makes their individual judgment less important and should have increased agreement.

Patient Classification Agreement and Implications for Payment

Next, we mapped each case into a CMG first using the FIM motor and cognitive scale scores and then using the pseudo-FIM motor and cognitive scale scores. The FIM scales and the pseudo-FIM scales from the MDS-PAC mapped into the same CMG 53 percent of the time. Several different approaches to improve the match between the mappings were subsequently tried. Ultimately, the best effort improved the level of agreement to 57 percent by using a regression mapping of pseudo-FIM items onto the FIM scores and by dropping one facility. The facility that we dropped had a mean difference in motor scores between the two instruments of 14 points (compared to an overall mean difference of 2.4). Further, that facility's team was only team to initially fail our certification exam.

To help understand whether agreement was better for some types of cases, we looked at agreement by RIC, the first tier within the payment system. CMG agreement within RICs was best for a few small RICs (which have only a few payment cells), and it was generally much lower among the larger RICs. Although this level of CMG agreement between instruments (53 to 57 percent) is low for use in a payment system, we found that scoring error within an instrument was high and led to equally poor levels of agreement, 50 percent for the FIM and 55 percent for the MDS-PAC (when the CMGs that result from calibration team responses are compared to institutional team responses on the same instrument).

Despite the poor levels of classification agreement, mean payment differences between the two instruments were small, averaging –$46, and not significantly different from zero. At the facility level, mean per case differences increased somewhat to $82. Despite good overall agreement, we found that more than 20 percent of the facilities would experience revenue differences of 10 percent or more. This remained true when we restricted our sample to hospitals with at least 50 cases. Our multivariate analysis of payment differences showed significant differences across hospitals but these were not systematically associated with patient or hospital characteristics.

Administrative Burden

By far the biggest difference between the instruments was their length. An important limitation of this study was that we did not examine the benefits of the expanded conceptual base provided by the MDS-PAC. We did, however, look at the costs in terms of the administrative burden.

Not unexpectedly, the administrative burden of the MDS-PAC overall was greater than that of the FIM. The magnitude of the difference was large, 147 minutes on average for institutional teams to complete the MDS-PAC compared to 25 minutes to complete the FIM, a sixfold difference. We found a clear learning curve effect during the study (average completion time for the first two weeks of the study of 184 minutes fell to 120 minutes for weeks 7 and 8), which could continue to reduce times beyond those reported here. The size of the data collection team also influenced data completion times significantly; the larger the team the longer the time. By the end of the study, one-person teams had times that were consistent with those reported in the November 3, 2000, Notice of Proposed Rule Making (85–90 minutes). Administration took longer for patients with lower motor function and for those with poor ability to communicate. Urban hospitals had lower times and there was notable variation across regions. The latter may be reflecting facility level differences that we did not control for.

In summary, our study's most important findings are (1) scoring reliabilities, while generally higher on the FIM than the PAC, were not as high as we would hope to see in an instrument intended for payment; (2) the best translation and mappings of the MDS-PAC into CMGs (created from FIM data) agreed with the FIM only 53–57 percent of the time; (3) despite this poor agreement, overall payment differences between the instruments were small; (4) however, 20 percent of the hospitals could see revenue differences of 10 percent or more depending upon which instrument is used; (5) all our multivariate analyses show strong random effects for hospitals with few other significant variables suggesting that additional training could help standardize responses and remove hospital-specific differences; and (6) the administrative burden associated with the MDS-PAC, 120 minutes compared to 23 minutes for the FIM at the end of the study, was substantial.

Instrument Specific Study Recommendations

If the MDS-PAC is selected as the basis of the instrument and the CMGs developed from the FIM are used, then we recommend the following:

- Add the list of impairment codes to the form and improve the guidance given for selecting the proper impairment code.
- Consider adding a scoring category between maximal assistance and total dependence that captures patients completing less than 25 percent of subtasks or change the definition of total dependence.

- Change or supplement the ADL Assist Codes—either add one-person torso and multiple limb or change one limb weight-bearing to one person.
- Revise the scoring to capture the distinction between independence and modified independence and collapse the setup and supervision categories.
- Identify wheelchair-dependent cases.
- Drop Metamucil® from the medication list.
- Continue to use medications to help distinguish complete independence from modified independence but drop medications from the appliance support list.
- Develop additional training materials to further standardize scoring.

In addition, the heavy administrative burden associated with this instrument is of concern. This suggests limiting the number of administrations and possibly limiting implementation to only those items that are relevant for rehabilitation. Items that are currently included on the MDS-PAC so that patient comparability across settings can be assessed might be deferred until the instrument is introduced in multiple settings.

If the FIM is selected, then we recommend enhancing the instrument by making explicit items that are implicitly being evaluated in the FIM scoring process. FIM scoring is deceptively complex and this should improve inter-rater reliabilities. For example, persons were misscored more than half the time when they were independent in eating but had chewing problems and/or swallowing problems that led to the use of modified diets. Similarly, in the locomotion item, FIM scores were not consistent with walking distances explicitly reported in the PAC. Thus, for the FIM, we would recommend the following:

- Standardize the assessment period.
- Add the list of impairment codes to the form and improve the guidance given for selecting the proper impairment code.
- Add explicit scoring aides to improve reliability including
  - Distance walked or traveled in a wheelchair,
  - Diet modification and chewing problems, and
  - Instructions to score locomotion item using expected mode at discharge.
- Separate and record both bowel continence and bowel management assistance.
- Separate and record both bladder continence and bladder management assistance.

- When scoring items such as transfer tub/shower where options are not equivalent, specify rules for which option is to be used and then record which option is being used.

Finally, we suggest that if this option is selected, consideration be given to creating a flexible add/drop section that allows for experimentation and the introduction of new items in the future.

Postscript

Policymakers elected to use a FIM-like instrument called the Patient Assessment Instrument (PAI). Study recommendations for instrument refinement, additional training, and scoring guidance were followed. A section for possible additional items has been added to the PAI and additional research is under way to evaluate the content and format of additional items.

Acknowledgments

This project was truly a multi-institutional collaborative effort conducted by faculty from Harvard Medical School, Sargent College of Health and Rehabilitation Sciences, and RAND. We are indebted to colleagues at the Hebrew Rehabilitation Center for the Aged and at Uniform Data Systems for Medical Rehabilitation for the training and consultative support they provided to the project. Both organizations were responsive and professional and gave generously of their time. This project would not have been feasible without their cooperation and assistance. We are also most grateful to the 50 rehabilitation hospitals that participated in the study and their staff. The study was on an extremely tight schedule and these institutions performed admirably. We want to express our appreciation, as well, to the New England hospitals that generously provided a training ground for our calibration teams. Rebecca Joyce, the project coordinator, orchestrated and handled the complicated project logistics with great competence. Daryl Caudry provided exceptional programming assistance during the analytic phase of the project. Cathy Sherbourne provided a thoughtful review that helped improve the report's clarity substantially. Finally, we would like to thank Grace Carter and Carolyn Rimes for their patience and constant support throughout this entire effort.

Acronyms

ADC	Average Daily Census
ADLs	Activities of Daily Living
BSCI	Brain or Spinal Cord Injury
CART	Classification and Regression Tree
CMAI	Cohen-Mansfield Agitation Inventory
CMG	Case Mix Group
CMS	Centers for Medicare and Medicaid Services (successor agency to the Health Care Financing Administration)
COG	Cognitive Scale
CPS	Cognitive Performance Scale
FAQ	Frequently Asked Question
FIM	Functional Independence Measure
FRG	Function Related Group
GDS	Global Deterioration Scale
HCFA	Health Care Financing Administration
ICC	Intraclass Correlation Coefficients
ICO	International Classification of Diseases
IRF	Inpatient Rehabilitation Facility
LOS	Length of Stay
MDS-COGs	Minimum Data Set—Cognition Scales
MDS-PAC	Minimum Data Set—Post-Acute Care
ML	Maximum Likelihood
MMT	Multiple Major Trauma

MOSES	Multidimensional Observation Scale for Elderly Subjects
NBSCI	No Brain or Spinal Cord Injury
PAC	Abbreviated form of MDS-PAC
PAI	Patient Assessment Instrument
PPS	Prospective Payment System
RIC	Rehabilitation Impairment Category
RUG	Resource Utilization Group
UDSmr	Uniform Data System for medical rehabilitation

1. Introduction

Background on the Development of a Prospective Payment System for Inpatient Rehabilitation

Inpatient rehabilitation was exempted from the Medicare Prospective Payment System (PPS) for hospital payment at its introduction in 1984. Before that time, hospitals were paid by Medicare on the basis of their historical costs, but as costs continued to increase, policymakers sought ways to limit cost growth. A PPS creates an incentive for cost containment by setting case mix adjusted prices in advance and limiting the amount of growth in future prices. The case mix adjusted prices are based on the *expected costs* of care for patients in each case mix class rather than on the *actual costs* of care delivered to a particular patient. Efficient hospitals keep the positive difference between the prospectively set payment and their actual costs of care and inefficient hospitals must absorb their losses. In a PPS, case mix adjustment is believed to be important, so that all patients maintain access to care. Without such adjustment, facilities might avoid higher-cost patients. Research at the time the hospital PPS was introduced demonstrated that diagnoses—the case mix basis of the Medicare short term acute hospital PPS—were not adequate to explain resource needs in the rehabilitation hospital population and that measures of functional status were needed to appropriately target payments to patient needs (Hosek et al., 1986). At that time, there was no agreement on what measures of functional status should be used, nor were these data routinely collected. Since then, the rehabilitation hospital community has developed a parsimonious, 18-item measure for this purpose, called the Functional Independence Measure (FIM) and has secured the voluntary participation of a substantial portion of inpatient rehabilitation providers in collecting these data. Stineman and her colleagues (1994a, 1997b) used the FIM to develop a patient classification system for medical rehabilitation, called the Functional Independence Measure—Function Related Groups (FIM-FRGs). A RAND team found these measures and methods to be a solid foundation for the preliminary design of a potential PPS for inpatient rehabilitation with a per case payment (Carter et al., 1997a, 1997b).

During this same period of time, research in another segment of the provider community—nursing facilities—was evolving along a separate path. In response to an Institute of Medicine Study in the mid 1980s calling for improvements in nursing home quality and more patient-centered care, researchers in this community developed a more comprehensive, multipurpose instrument called the Resident Assessment Instrument—Minimum Data Set (MDS). This instrument is mandated for use in all nursing facilities and is used for care planning, patient classification, and quality assurance. The patient classification system, Resource Utilization Groups III (RUGs III), was implemented from the MDS and went into effect in 1998 (see Fries et al., 1994). The RUGs system uses a per diem payment and is the foundation of the nursing facility PPS.

Since the introduction of the hospital PPS, hospital length of stay has fallen dramatically whereas discharges to all types of post-acute care providers (rehabilitation hospitals, nursing facilities, and home health agencies) have increased markedly. In an effort to control costs in the post-acute care area, the Balanced Budget Act of 1997 mandated the introduction of prospective payment systems for nursing facilities, rehabilitation hospitals, and home health agencies. With this growth in the use of post-acute care providers has come increased recognition of considerable overlap in populations being treated in each setting. Many nursing facilities now specialize in subacute and rehabilitation care or have special units within them to attract these patients. Thus, policymakers have called for a more integrated approach to patient assessment that will cross post-acute settings. The Minimum Data Set—Post-Acute Care (MDS-PAC) was developed as a response to this need for integration across settings. The desire to implement a prospective payment system in the near future led MDS-PAC designers to include elements in the MDS-PAC that would allow one to map from it to the FIM scales used in the original payment system design. A consequence of this integration effort is that the MDS-PAC contains many items that do not exactly replicate those in either the FIM or the MDS.

The FIM is an 18-item measure that was constructed as a minimal instrument to evaluate and monitor functional and cognitive status in inpatient rehabilitation settings. Each item is rated on a seven-point scale from total dependence (1) to total independence (7). The FIM is often described as having two domains, a motor score domain (items 42A–M) and a cognitive score domain (items 42N–R). See Appendix A for a copy of the instrument.

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

Background on Instrument Performance

Scaling and assessment instruments are usually evaluated on a number of dimensions including validity, reliability, and internal consistency. The validity of an instrument refers to the extent to which the instrument actually measures the concepts that it intends to measure. Face validity is often judged by professionals after reviewing item and scale content. Validity may also be established by demonstrating that scales and constructs within an instrument have the expected empirical relationships to external measures or within subgroup structures. The reliability of an instrument refers to its performance under repeated measurement either by different evaluators or possibly by the same evaluator but at different times. A reliable instrument will produce very similar estimates within close time frames from the same or different assessors. Reliability is usually reported by measures of association such as intraclass and/or Pearson correlations, or with measures of agreement such as (weighted or unweighted) kappa coefficients, which also correct for chance agreement. The internal consistency of a scale refers to the intercorrelations among its items, and scales with higher internal consistency are thought to have higher test-retest reliability. Internal consistency is usually measured using Cronbach's coefficient alpha.

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

The inter-rater reliability of the FIM has been assessed in several studies. In an early study of 89 facilities, unweighted item level kappa coefficients ranged from .53 (moderate agreement) to .66 (good agreement). For the subset of facilities that had passed a competence exam, scores were notably higher ranging from .69 (good agreement) to .84 (excellent agreement). Intraclass correlation coefficients (ICC) for the motor domain were .96 and .91 for the cognitive domain (Hamilton et al., 1994). Test-retest reliability was assessed on 45 cases by Pollak et al. (1996), who found motor score ICC = .9 and cognitive score ICC = .8. Sharrack et al. (1999) found that inter-rater agreement varied with kappa coefficients ranging from .26 (poor agreement) to .88 (excellent agreement); ICC ranged from .56 to .99. Segal et al. (1993) found that although the total reliability score was good (.83), reliability coefficients across individual items varied markedly from .02 (poor agreement) to .77 (very good agreement).

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

Pilot studies of an earlier version of the MDS-PAC have been undertaken but results have not been reported in the peer-reviewed literature. One pilot study of the time to complete the MDS-PAC in rehabilitation hospitals reported 105 minutes for the first few assessments dropping to 85 minutes after 10 or more cases. Similar numbers for nursing home staff, who probably benefited from familiarity with completing the MDS, were 85 and 77 minutes. The pilot results of an inter-rater reliability study of 171 cases that found that average reliability of 315 MDS-PAC items on draft nine was .78 with a range of .51 to 1.00 (Health Care Financing Administration (HCFA), 2000). Since the MDS-PAC was developed from the MDS for nursing facility residents, which has undergone considerable testing, some of those findings are reported here.

Snowden et al. (1999) examined the construct validity of the MDS cognitive, Activities of Daily Living (ADLs),[1] and behavior domains comparing them to the Folstein Mini-Mental Status Exam, the Dementia Rating Scale scores, and the

---

[1] Eating, walking, grooming, bathing, toileting, and dressing.

Alzheimers Disease Patient Registry physician behavior checklist and concluded that the MDS data demonstrate reasonable criterion validity for research purposes.

Casten et al. (1998) used confirmatory factor analysis on MDS data to evaluate five domains within the MDS: cognition, activities of daily living, time use, social quality, depression, and problem behaviors. For cognitively intact individuals and all residents together, the domain clusters except social quality were confirmed. For individuals with serious cognitive impairment, none of the domains were confirmed.

Lawton et al. (1998) provided construct validity by testing the confirmed MDS domains (ADLs, cognition, time use, depression, and problem behaviors) against established clinical research measures such as the Blessed Test, Reisberg Global Deterioration Scale (GDS), ADLs, Geriatric Depression Scale, Raskin Depression, Positive Affect, Negative Affect, Mattis Total, Multidimensional Observation Scale for Elderly Subjects (MOSES) Irritability, MOSES Depression, and Cohen-Mansfield Agitation Inventory. They found that the majority of their hypotheses were confirmed but that validity coefficients were modest and performance for depression and problem behaviors was not as good as for ADLs, cognition, and time use.

In a multistate evaluation of the MDS, researchers found that items in key areas of functional status (cognition, ADLs, continence, and diagnoses) had ICCs of .7 or higher, that 63 percent of the items had ICCs of .6 or higher, and that 89 percent had ICCs of .4 or higher (Hawes et al., 1995). That instrument has been translated and used in 15 other countries and has undergone reliability testing in six (Hawes et al., 1997; Sgadari et al., 1997).

The cognitive items on the MDS and the MDS-PAC are thought to be particularly promising and have led to the development of two different scales. The MDS cognitive performance scale (CPS) (see Morris et al., 1994) has been validated against the Mini-Mental Status Exam (Hartmaier et al., 1995) and the MDS cognition scale (MDS-COGs) against the GDS and the Mini-Mental Status Exam (Hartmaier et al., 1994). With minimal recoding, the CPS can be scaled from the MDS-PAC. The MDS-COGs measure must be rescaled for the MDS-PAC, as the MDS-PAC contains a different and smaller subset of memory/recall ability items.

Purpose and Scope of This Project

The purpose of this project is to evaluate the MDS-PAC for use in classifying cases into case mix groups (CMGs) in the planned inpatient rehabilitation prospective payment system. The MDS-PAC integrates elements from the nursing facility Resident Assessment Instrument—Minimum Data Set with concepts from the rehabilitation hospital's assessment tool, the Functional Independence Measure. Although both of the underlying instruments have been tested and an earlier version of this instrument has undergone some fieldwork, the final instrument has not been field-tested. To this end, we will evaluate proposed strategies (and investigate new ones) for mapping MDS-PAC data into FIM data and then into CMGs and evaluate the psychometric properties of the MDS-PAC as compared to the FIM.

Organization of the Report

Section 2 describes the study design and implementation, which includes the recruitment and enrollment of facilities. It also includes data on the characteristics of our final sample of hospitals and covers training and certification procedures for the institutional data collectors and the study calibration teams. Section 3, the first phase of our analytic work, describes how we use items from the MDS-PAC to create "pseudo-FIM" items. Some basic comparisons of the two instruments are given in Section 4. Section 5 provides data on how well the pseudo-FIM items from the MDS-PAC compare to actual FIM items. Case mix classification and payment analyses are presented in Section 6.

All references to appendices in this report are references to the companion volume, J. L. Buchanan, P. Andres, S. M. Haley, S. M. Paddock, D. C. Young, and A. Zaslavsky, *Final Report on Assessment Instruments for a Prospective Payment System: Appendices,* Santa Monica, CA: RAND, MR-1501/1-CMS, 2002.

2. Study Design and Implementation

This section of the report describes the study design, implementation procedures including facility recruitment and selection, and the data collection training and oversight methods.

Research Questions

This study was designed to address the following questions:

1. How accurate is the MDS-PAC for use in classifying cases into CMGs for the proposed inpatient rehabilitation prospective payment system?
2. How do the validity, reliability, and consistency of the FIM and the MDS-PAC compare?
3. What are the time costs associated with data collection on each instrument?

Sample Size

We use a two-tiered study design. The first tier provides our primary analytic samples, and the second tier is needed to address some of the psychometric issues we are concerned about. For tier one, we recruited 50 institutions and each participating institution was trained on the MDS-PAC and the study. They were then asked to complete both the FIM and the MDS-PAC on all new Medicare admissions with stays beyond three days for an eight-week period.

The second tier of our design is intended to ensure that institutions across the country are all calibrating to the same set of norms to give us approximate measures of inter-rater reliability. For this component, we hired three "calibration teams," which visited each participating institution. These teams spent one to two days at each institution and rated an average of four cases using both the FIM and the MDS-PAC. The strength of this strategy is that it allowed us to measure how well calibrated facilities were and whether there were any specific regional differences for a substantial number of facilities. A disadvantage may be that the assessment processes used by the calibration team may differ from those used by the institutional providers and thus our measures of inter-rater reliability could be less precise. Because one set of ratings (those done by the institutional providers) is part of a patient care process and the other (those done by the calibration team) is not, the two assessment processes can never be identical and consequently we will always have imperfect measures of inter-rater reliability. However, we were most interested in how one instrument performed relative to the other. So as long as the deviation in the calibration team rating process did not systematically disadvantage one instrument over the other, this procedure should be adequate.

The institutional sample is shown in Row 1 and the calibration sample in Row 2 of Table 2.1. Both samples were used for the psychometric and instrument performance comparisons described in Section 4. The institutional assessment sample (Row 1) was the primary sample used for the classification and agreement analyses in Sections 5 and 6.

Table 2.1

Sample Size

	Study Tier	Number of Institutions	Completion Time Data	FIM	MDS-PAC	Number of Cases
Institutional assessments	1	50	Yes	Yes	Yes	3,484
Calibration assessments	2	50	Yes	Yes	Yes	241

Hospital Recruitment

The study was conducted in "FIM-certified" inpatient rehabilitation facilities (IRFs). Uniform Data System for Medical Rehabilitation (UDSmr), the organization that trains, certifies, and collects FIM data, reported over 650 FIM-certified IRFs at the time of the study. The strategy of using UDSmr-certified facilities had several advantages. First, all participants were trained and able to perform FIM assessments in a standardized manner and they were able to collect and report regular assessment data on all patients. Second, the study had a method of identifying FIM-certified IRFs and communicating with them rapidly. Third, these facilities included hospitals that cared for over two-thirds of all Medicare cases. (Freestanding and large facilities are overrepresented.) The limitation to this strategy was that all participants were much more familiar with the FIM than with the MDS-PAC.

An invitation packet containing a letter from Harvard Medical School describing the study, a letter from the Health Care Financing Administration endorsing the study and encouraging participation, and a response form (see Appendix C) was sent out to approximately 650 FIM-certified IRFs over the UDSmr fax broadcast network in early August 2000. The invitation described the study and indicated that participating facilities would be expected to collect MDS-PAC data (in addition to the FIM data already being collected) on all Medicare admissions for two months. For each completed MDS-PAC, FIM pair that the study received, the IRF would receive a $35 payment up to a $4,000 maximum per IRF.

Participating IRFs were expected to send one or more data collection teams to a two-day training session on the MDS-PAC. IRFs with up to 50 Medicare admissions a month were required to send only one team; IRFs with 51–100 Medicare admissions a month were asked to send two teams, and those with more than 100 Medicare admissions a month were to send three teams. The study paid $500 to each team member attending the training. The payment went to the participant if the course was completed on off-duty time and to the IRF otherwise. The data collection teams were multidisciplinary and consisted of two to four clinicians, including nurses, physical and occupational therapists, speech language pathologists, and a small number of other professionals (e.g., social workers, recreational therapists, and psychologists). One week after the broadcast was completed, approximately 170 response forms had been received from facilities indicating a desire to participate. The response forms included information on facility characteristics

Facility Selection

Several criteria were used in selecting facilities for inclusion in the study from the 170 applicants. First, facilities from both urban and rural areas were needed, as were both hospital-based and freestanding facilities. Next, a representative distribution by size was needed. Small facilities included those with average daily census ≤ 20 patients, medium included those with average daily census between 21 and 50 patients, and large facilities were those with an average daily census > 50 patients. Finally, the set of selected hospitals was to be geographically clustered but representative of the country as a whole.

Fifty-three facilities were originally selected for study participation; two dropped out before facility data collectors could be trained, an additional facility dropped out after the training because of problems in timing and responsiveness of their Institutional Review Board, leaving 50 study IRFs. Eight of the 50 facilities (16 percent) were classified as rural and 14 facilities (28 percent) as freestanding (see Table 2.2). The facilities are distributed across the country and include 22 states. A map of the sites is shown in Figure 2.1. Additional details on the distribution of facilities by size and geographic area is given in Appendix D. Data on nonrespondents were not available.

Table 2.2

Characteristics of Selected Facilities

	Hospital-Based	Freestanding	All Facilities
Type	71%	28%	100%
Size—mean ADC[a]	25	64	36
% small (ADC ≤ 20)	42%	7%	31%
% medium (21–50 ADC)	55%	33%	49%
% large (ADC > 50)	3%	60%	20%
Medicare admissions (per month)			
Mean	39	69	48
Median	34	60	37
Minimum	10	20	10
Maximum	112	148	148
Rural	22%	0%	16%

[a]Average daily census.

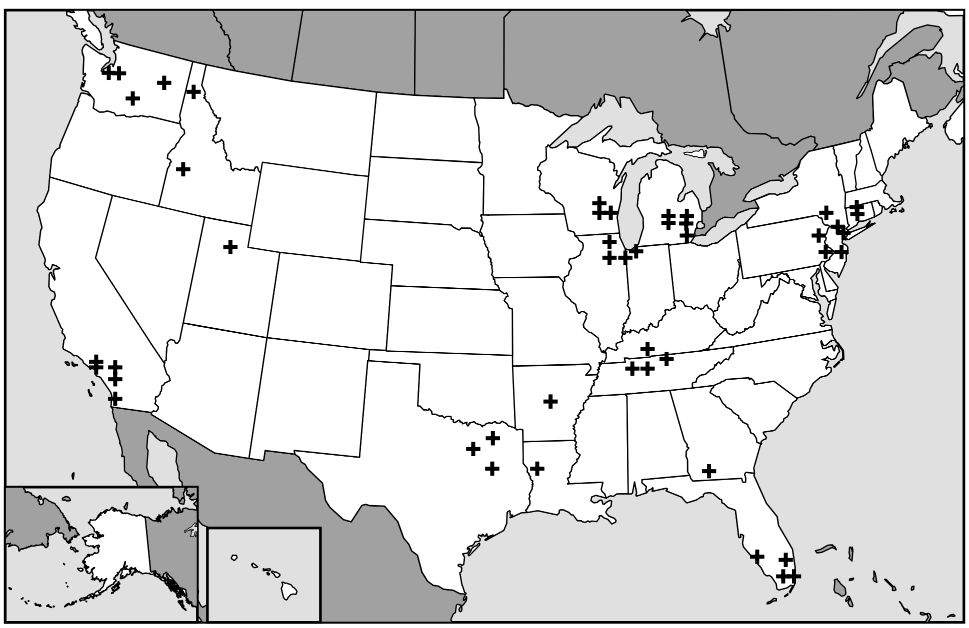

Figure 2.1—Map of Selected Facilities

Training the Trainers

With 51 facilities, approximately 70 data collection teams and up to 280 individuals were expected to train on the MDS-PAC. The study hired faculty from MDS-PAC development team to provide a three-day "train the trainers" program on the MDS-PAC. The trainers trained in this program were instructors from UDSmr. The course was an expansion of earlier MDS-PAC training sessions that had been provided as the instrument was being developed and field-tested. The session included an item-by-item review, a video with several segments to be scored, a visit to a local facility to score actual cases, and a final debriefing on the field experience. An additional segment of the course, run by the Field Core Coordinator, taught participants how to complete the study tracking forms.

Each participant received a binder with (1) a written manual on the MDS-PAC, (2) a short description of the project and its goals, (3) study forms and instructions, and (4) instructions for following the study protocol within each IRF. As part of the final certification procedure, each participant was required to return to their home facility and complete two MDS-PAC assessments and to copy the MDS-PAC forms and return them to the Field Core Coordinator. The course trainers reviewed the two MDS-PAC forms and then completed an hour-long telephone debriefing. During this call, study researchers corrected form completion errors and asked clarifying questions about the practice administrations. The review culminated in administration of an oral "MDS-PAC certification exam" designed to review technical and content mastery of the instrument. If there were too many errors or sessions went poorly, participants were asked to complete another set of cases and repeat the debriefing. All the study trainers passed the certification test on the first round.

The UDSmr trainers had all undergone some MDS-PAC training before this session and many had already completed 20 or more MDS-PAC cases. Consequently, they had substantive questions and some important thoughts on how to strengthen the training session. The trainers felt strongly that they needed more scoring exercises to help ensure that they were all rating patients in the same manner on functional assessment items. In response to this, we prepared a detailed case study (with a written rationale for each score) that each trainer successfully completed. We also added short one-paragraph vignettes with scoring rationale to the training slides following most of the functional status items. Further, we modified the certification process for IRF trainees to include a case study.

The trainers were sent a new set of training materials, which included a videotape of scoring examples, a videotape of the Field Core Coordinator explaining the study protocol and use of study-specific tracking forms, a set of overhead transparencies that included short one-paragraph vignettes to score following the functional assessment items, and a new project manual with (1) a revised MDS-PAC manual, (2) the project description and history, (3) study forms, (4) (somewhat modified) instructions for following the study protocol within each IRF, and (5) a new description of the homework and certification process for the trainees.

Facility Training

Once the 50 IRFs were enrolled in the study, we identified 10 geographic clusters for training. One two-day training session was held in each of these sites (Stamford, CT; Philadelphia, PA; Miami, FL; Detroit, MI; Chicago, IL; Nashville, TN; Milwaukee, WI; Dallas, TX; Spokane, WA; and Los Angeles, CA). Nine of the 10 training sessions were completed in the last two weeks of August and the tenth session (Los Angeles) was held on the weekend of September 9 and 10. After the first day of each training session, the trainer had a conference call with the Field Core Coordinator and other study investigators to report on how the session was progressing and to obtain answers to any questions that arose during the course of the session for which trainer felt uncertain.

Each individual attending the training session received a project manual (binder) with (1) a description of the project and project history, (2) a MDS-PAC manual, (3) study forms, (4) instructions for following the study protocol, and (5) a description of the homework and certification process. All but three or four teams completed the certification process within one week of training. For those who completed the training and certification process, data collection began on September 5. In all, 69 teams, which included 262 individuals, were trained.

Each IRF data collection team left the MDS-PAC training program with "homework and certification" instructions to (1) complete two practice MDS-PAC administrations (one of which would be copied and sent to the Field Core Coordinator), (2) score a written case study designed to highlight scoring and interpretation issues on the MDS-PAC, and (3) complete a "study protocol exam" highlighting key issues on data collection, use of study forms, and overall study protocol. All homework and certification materials were reviewed by the Field Core Coordinator and Field Core Designer and graded for "technical" accuracy (appropriate use of the MDS-PAC form) and content (case study, protocol exam). Deficiencies were noted and immediately communicated (via fax) to the institutional team using a feedback and certification form. Several teams received additional counseling on data collection via telephone conference call. All 69 teams were certified for data collection by early September.

Data Collection and Transmission

Before the first week of data collection, each participating IRF was sent a box containing 50 pre-labeled sets of study forms (FIM face sheets and forms, MDS-PAC face sheets and forms), workflow tracking forms, and pre-addressed FedEx mailers for returning forms. Additional boxes of pre-labeled data collection forms were sent, as needed, to larger facilities. Each study form had a study ID label that identified the IRF uniquely and patients sequentially. This allowed us to account for all study forms and to match FIM and MDS-PAC forms for the same patient without bringing any identifiable data to Harvard.

The admission and workflow tracking form (see Appendix E) was used by the IRF to record all admissions and to identify those eligible for study admission (Medicare-eligible and staying at least three days). This form was maintained by the IRF and is the link that identifies study patients with their study assigned ID. It also functioned as a workflow tracking tool that helped the IRF team coordinator check that all tasks were completed on all study-eligible admissions.

The FIM and MDS-PAC face sheets (see Appendix E) were used to collect information about when the assessments were done, who did them, and how long they took to complete. Data collectors were instructed to fax these to Harvard on a toll-free line each day.

The actual FIM and MDS-PAC forms used by the study are shown in Appendixes A and B. These were collected within the IRF and sent via FedEx (using the pre-paid, pre-addressed forms) to Harvard biweekly.

Monitoring and Communication with the Field

The Field Core Coordinator had a single point of contact at each IRF and maintained regular communication with 50 data collection sites. This communication was intended to monitor productivity and to facilitate calibration team visits. In addition, the study maintained a toll free data collection hotline that all data collectors used to get clarification on MDS-PAC scoring issues or to discuss questions on study protocol. These calls were directed to the Field Core Coordinator who either answered them directly or triaged to one of the instrument development teams. All questions and answers on the MDS-PAC were included in a hypertext-linked document, the MDS-PAC FAQs (frequently asked questions) (see Appendix F). During the data collection phase of the study, the FAQ document was posted, along with other study-related documents and communication, on a web site designed to facilitate communication between the Field Core Coordinator, trainers, and IRF data collectors.

Newsletters were another form of communication used with study field staff. These contained updates on study changes and on study progress. Copies of the newsletters are contained in Appendix G. When a potentially serious problem arose with regard to data collection protocol, a Study Protocol Update (see Appendix G) was issued to alert all data collectors and inform them of proper procedures to follow.

Calibration Teams

Nine individuals were hired by the study to form three calibration teams. Each team was to have a nurse and two therapists. The therapists included three physical therapists, two occupational therapists, and one speech language pathologist. Each team was initially configured with a nurse, a physical therapist, and either an occupational therapist or a speech language pathologist. Two of the nurses had family emergencies, one in the middle of training and one three weeks into data collection, and could not be replaced. Data collection continued with two teams of two and one team of three.

All calibration team members underwent four weeks of intensive training in Boston. During the first week, they attended a three-day training and certification session on the MDS-PAC similar to that given earlier to the trainers. They also underwent a one-day training and certification session on the FIM given by the Director of Training from UDSmr. This was followed by three weeks of practice in four settings. The settings differed significantly in organization and structure (see Appendix H for additional detail). That exposure to different settings gave the teams the opportunity to see significant organizational differences and also gave them experience entering new facilities and orienting themselves for immediate data collection. During training, each calibration team member completed between 25 and 30 practice MDS-PACs and FIMs. The final composition of the three calibration teams for field deployment was designated during the third week of training and, thereafter, they practiced as a team to refine data capture techniques and efficiency.

Competence on the MDS-PAC and study protocol was assessed at a number of points during and after calibration team training. These procedures included (1) oral review and question-and-answer sessions at numerous points during training, (2) completion of a written case study at week 2, (3) completion of the data collection protocol exam and MDS-PAC certification exam at week 3 (same as provided to UDSmr trainers), (4) assessment of inter-team reliability (concurrent MDS-PAC and FIM assessments performed by calibration team pairs) at week 4, and (5) completion of a final paper case study to assess inter-team reliability in scoring the MDS-PAC at week 7.

During the data collection phase of the study, at least one of the calibration teams traveled to each of the 50 sites in the study. Team assignments were made in a way to ensure that each team covered the country maximally. The travel schedule was organized around the 10 training sites, scheduling two and sometimes three visits to most sites. At least two teams visited each geographic cluster and all teams visited a cluster when three visits were necessary. Each team visited sites on both coasts and in the middle of the country. All teams visited sites in the northern and southern part of the country. For small facilities, we were not always able to get an appropriate study sample in one visit, so revisits were scheduled as needed.

In large facilities, more cases were occasionally available than could be scored, so a formal set of sampling instructions were developed for the calibration team. These are provided in Appendix I. We also wanted to determine whether the order of form completion between the FIM and the MDS-PAC made any difference, so our sampling instructions indicated the order of form completion by the calibration team. For the cost analysis, there was some concern that the first team to complete the MDS-PAC would require more effort, so we randomized the assignment order between the facility and the calibration teams. These assignments are all embodied in the sampling instructions in Appendix I.

Because the timing of FIM data collection was not well standardized, we asked the calibration teams to code both the admission scores and the reference day (day 3) scores on the same FIM form. The latter were recorded in the coding positions for discharge FIM scores on the FIM form. This revision gave us a direct measure for each patient of the calibration team assessments of how FIM scores might have changed.

The calibration teams were asked to review the Admission and Workflow Tracking forms at each IRF visited to ensure that the facility teams were handling these procedures correctly. As part of this review, they provided us with some summary statistics on facility performance at that time. In addition, they provided feedback on their impressions of how strictly the IRF was adhering to the study protocol.

The calibration teams maintained weekly or more frequent contact with the Field Core Coordinator to balance assignments and to backfill when cases did not occur as anticipated. The completed calibration forms were mailed to Harvard biweekly using pre-addressed, prepaid forms.

3. Translating the MDS-PAC into FIM Motor and Cognitive Scale Items

To understand the implications of substituting the MDS-PAC for the FIM—a method for using information from the MDS-PAC to classify patients into CMGs—the patient classification system for rehabilitation hospitals was needed. The CMG classification system placed patients into a rehabilitation impairment category (e.g., stroke, traumatic brain injury, or spinal cord injury), gave the underlying reason the patient is in a rehabilitation hospital, and then placed the patient into a class within the selected rehabilitation impairment category on the basis of patient age, FIM motor scale score, and FIM cognitive scale score.

The FIM motor scale score is the sum of the 13 individual motor item scores and the FIM cognitive scale score is the sum of the five individual cognitive items. A copy of the very simple scoring template for the 18 FIM items is shown in Figure 3.1. Scores range from 1 for total dependence to 7 for total independence and actually have fairly complex scoring rules that differ somewhat by item. For example, the locomotion item has an explicit distance requirement that is not signaled on the actual scoring sheet but is imbedded in the written scoring rules. As part of the FIM training, UDSmr provides a detailed training manual with decision-tree-like scoring instructions for the different levels of each item. Additional training materials, called FIM lessons, are also available to help therapists learn the scoring nuances. The FIM is a measure of disability and burden of care. Safety and the time required to complete an activity also influence scoring. The FIM was designed to be used by trained clinicians but was intended to be discipline-free. All 18 items must be completed so any activity that cannot be completed is scored as 1, total dependence. Admission scores must be completed during the first 72 hours after admission but generally refer to performance over the past 24 hours. Scoring instructions indicate that the best available information should be used and that direct observation of subject performance is preferred. At the time of this study, roughly 60 percent of the industry voluntarily used the FIM and submitted their data to UDSmr. Other institutions used it without formal certification or participation in UDSmr data collection. In many institutions, FIM language on levels of assistance has become a standard way for therapists to communicate with one another about patient performance.[1]

---

[1] Item-by-item scoring rules are available at www.hcfa.gov/medicare/irfpai-manual.htm.

42. FIM		FIM Levels
Self-care		**No helper**
A. Eating		7 Complete independence (timely, safely)
B. Grooming		6 Modified independence (device)
C. Bathing		**Helper—modified dependence**
D. Dressing upper body		5 Supervision (subject = 100%)
E. Dressing lower body		4 Minimal assistance (subject = 75% or more)
F. Toileting, sphincter control		3 Moderate assistance (subject = 50% or more)
G. Bladder management		**Helper—complete dependence**
H. Bowel management		2 Maximal assistance (subject = 25% or more)
Transfers		1 Total assistance (subject < 25%); not testable
I. Bed, chair, wheelchair		
J. Toilet		
K. Tub/shower		
Locomotion		
L. Walk/wheelchair		
M. Stairs		
Communication		
N. Comprehension		
O. Expression		
Social cognition		
P. Social interaction		
Q. Problem solving		
R. Memory		

Figure 3.1—FIM Scoring

With its origins in the nursing home minimum data set, the MDS-PAC differed substantially from the FIM in both the breadth of coverage and its approach to assessment. The MDS-PAC was viewed as a multipurpose information gathering tool and data collectors were instructed to consult the patient, the patient's family, and all caregivers from all shifts during the first three days of the patient's hospital stay, as well as to review the chart. Another difference between the instruments was that the FIM often instructed scorers to use the most dependent episode, whereas the MDS-PAC scorers were instructed to collect data over this longer time frame and to use a more comprehensive consultation list but to allow one or two more dependent episodes before scoring patients to a more dependent level. Like the FIM, the MDS-PAC is scored on a seven-point scale, but scoring is from 0 to 6 and uses the reverse orientation, so in the MDS-PAC 0 represents total independence (the parallel of FIM 7), and 6 represents total dependence (FIM 1). As a relatively long instrument, the MDS-PAC relies more on written instructions and multiple items for completing the form. An example of this is the treatment of physical assistance in the performance of self-care activities. In the FIM, the amount of physical assistance provided influences the level of dependence scored. In contrast, the MDS-PAC first scores the level of self-performance and then records the amount of physical assistance received in another item (see Figures 3.2 and 3.3). Thus, to use the MDS-PAC information to create FIM motor and cognitive scale scores, rules for combining MDS-PAC elements into each of the 18 FIM items were needed.

Before finalizing the MDS-PAC, several conversations took place between the two instrument development teams that revised some MDS-PAC items and added others with the intention of improving the ability to perform "FIM-like" scoring.

The Morris Translation

As part of the MDS-PAC instrument development effort, Dr. John Morris prepared the first set of scoring rules that combined multiple MDS-PAC items to translate them into each of the 18 FIM items (see Appendix K). This translation is particularly important for the five cognitive items as there is no direct correspondence between these items in the two instruments. We began our work using the Morris translation and found that although the translation worked fairly well for the cognitive items (mean cognitive score from the FIM was 28.50 compared to 28.51 using the Morris translation of the MDS-PAC into pseudo-FIM items), it performed less well on the motor items (mean motor score from the FIM 45.46 compared to 50.26 using the Morris translation of the MDS-PAC into pseudo-FIM motor items). Table 3.1 shows the scale and item mean comparisons for the two instruments. Although the overall motor scale difference is nearly five points, individual item means are reasonably close (within .6 point) except for the locomotion item where the mean difference is more than 1.5 points. On 12 of the 13 items, the pseudo-FIM means exceed the FIM means. The bowel management item is the exception with a larger FIM mean.

This substantial difference in the motor scale means led us to review the translation rules and the instrument scoring rules for both instruments. As several members of our research team were clinicians and had attended both the MDS-PAC and the FIM training sessions, we also had notes from these sessions to guide us. We soon realized that there were scoring rule differences that had not been accounted for in the Morris translations. This led to the development of a second set of translations for the motor scale items. The translation also benefited from further consultation with clinicians at UDSmr. In making clinical judgment about rescoring items, particularly the bowel and bladder items, where agreement was poor and scoring on multiple components could be considered inconsistent, we focused on what was most likely occurring *at*

SECTION E. FUNCTIONAL STATUS

1. **3 DAY ADL SELF-PERFORMANCE—**(***CODE for Performance Over All Shifts, for All Episodes, OVER LAST 3 DAYS***) [NOTE - for *Bathing and Tub Transfer,* code for most dependent single episode in this period]

0. *INDEPENDENT*—No help, setup, or supervision —OR— Help, setup, or supervision provided only 1 or 2 times during period (with any task or subtask)

1. *SETUP HELP ONLY*—Article or device provided or placed within reach of patient 3 or more times

2. *SUPERVISION*—Oversight, encouragement or cuing provided 3 or more times during period —OR— Supervision (1 or more times) plus physical assistance provided only 1 or 2 times during period (for a total of 3 or more episodes of help or supervision)

3. *MINIMAL ASSISTANCE (LIMITED ASSISTANCE)*—Patient highly involved in activity; received physical help in guided maneuvering of limbs or other non-weight bearing assistance 3 or more times —OR— Combination of non-weight bearing help with more help provided only 1 or 2 times during period (for a total of 3 or more episodes of physical help)

4. *MODERATE ASSISTANCE (EXTENSIVE ASSISTANCE)*—Patient performed part of activity on own (50% or more of subtasks) *BUT* help of following type(s) provided 3 or more times:
— Weight-bearing support (e.g., holding weight of limb, trunk)
— Full staff performance of a task (some of time) or discrete subtask

5. *MAXIMAL ASSISTANCE*—Patient involved but completed less than 50% of subtasks on own (includes 2+ person assist), received weight bearing help or full performance of certain subtasks 3 or more times

6. *TOTAL ASSISTANCE (TOTAL DEPENDENCE)*—Full staff performance of activity during entire period

8. *ACTIVITY DID NOT OCCUR*—During entire period

a. **BED MOBILITY**—How patient moves to and from lying position, turns side to side, and positions body while in bed	
b. **TRANSFER BED/CHAIR**—How patient moves between surfaces—to or from: bed, chair, wheelchair, standing position (**EXCLUDE** to/from bath/toilet)	
c. **LOCOMOTION**—How patient moves between locations in his/her room and adjacent corridor on the same floor. If in wheelchair, how moves once in wheelchair	
d. **WALK IN FACILITY**—How patient walks in room, corridor, or other place in facility	
e. **DRESSING UPPER BODY**—How patient dresses and undresses (**street clothes, underwear**) above the waist, includes prostheses, orthotics, fasteners, pullovers, etc.	
f. **DRESSING LOWER BODY**—How patient dresses and undresses (**street clothes, underwear**) from the waist down, includes protheses, orthotics, belts, pants, skirts, shoes, and fasteners	
g. **EATING**—How patient eats and drinks (regardless of skill), includes intake of nourishment by other means (e.g., tube feeding, total parenteral nutrition)	
h. **TOILET USE**—How patient uses the toilet room (or commode, bedpan, urinal); cleanses self after toilet use or incontinent episode(s), changes pad, manages ostomy or catheter, adjusts clothes (**EXCLUDE** transfer toilet)	
i. **TRANSFER TOILET**—How patient moves on and off toilet or commode	
j. **GROOMING/PERSONAL HYGIENE**—How patient maintains personal hygiene, including combing hair, brushing teeth, shaving, applying makeup, washing/drying face and hands (**EXCLUDE** baths and showers)	
k. **BATHING**—How patient takes full-body bath/shower or sponge bath (**EXCLUDE** washing of back and hair and TRANSFER). Includes how each part of body is bathed: arms, upper and lower legs, chest, abdomen, perineal area. ***Code for most dependent episode***	
l. **TRANSFER TUB/SHOWER**—How patient transfers in/out of tub/shower ***Code for most dependent episode***	

2. **ADL ASSIST CODES** (**Code for most help in last 3 days**)

0. Neither code applies
1. Weight bearing support with 1 limb
2. 2+ person physical assist

a. Bed mobility		g. Eating	
b. Transfer bed/chair		h. Toilet use	
c. Locomotion		i. Transfer	
d. Walk in facility		j. Grooming/personal hygiene	
e. Dressing upper body		k. Bathing	
f. Dressing lower body		l. Transfer tub/shower	

Figure 3.2—MDS-PAC Section E: Functional Status

SECTION F. BLADDER/BOWEL MANAGEMENT

1.	BLADDER CONTINENCE (Code for last 7-14 days)	a. *Control of urinary bladder function (if dribbles, volume insufficient to soak through undergarments)* 0. *CONTINENT*—Complete control; DOES NOT USE any type of catheter or other urinary collection device 1. *CONTINENT WITH CATHETER*—Complete control with use of any type of catheter or urinary collection device that does not leak urine 2. *BIWEEKLY INCONTINENCE*—Incontinent episodes less than once a week (i.e., once in last 2 weeks) 3. *WEEKLY INCONTINENCE*—Incontinent episodes once a week 4. *OCCASIONALLY INCONTINENT*—Incontinent episodes 2 or more times a week but not daily 5. *FREQUENTLY INCONTINENT*—Tended to be incontinent daily, but some control present (i.e., on day shift) 6. *INCONTINENT*—Has inadequate control of bladder, multiple daily episodes all or almost all of time 8. *DID NOT OCCUR*—No urine output from bladder b. *Is now more impaired in bladder continence than was prior to precipitating event (item **A7a**)* 0. No or unsure 1. Yes, more impaired today	
2.	BLADDER APPLIANCE (Code for last 24 hours)	*CODE:* 0. No 1. Yes a. External catheter b. Indwelling catheter c. Intermittent catheterization d. Medications for control e. Ostomy f. Pads, briefs g. Urinal, bedpan	
3.	BLADDER APPLIANCE SUPPORT (Code for last 24 hours)	0. No appliances **(in item F2)** 1. Use of appliances, did not require help or supervision 2. Use of appliances, required supervision or setup 3. Minimal contact assistance (light touch only) 4. Moderate assistance; patient able to do 50% or more of sub-tasks involved in using equipment 5. Maximal assistance; patient able to do 25-49% of all sub-tasks involved in using the equipment 6. Total dependence	
4.	BOWEL CONTINENCE (Code for last 7-14 days)	0. *CONTINENT*—Complete control, does not use ostomy device 1. *CONTINENT WITH OSTOMY*—Complete control with use of an ostomy device that does not leak stool 2. *BIWEEKLY INCONTINENCE*—Incontinent episodes less than once a week (i.e., once in last 2 weeks) 3. *WEEKLY INCONTINENCE*—Incontinent episodes once a week 4. *OCCASIONALLY INCONTINENT*—2-3 times a week 5. *FREQUENTLY INCONTINENT*—4+ times a week but not all of time 6. *INCONTINENT*—All of time 8. *DID NOT OCCUR*— No bowel movement during the entire 14 day assessment period	
5.	BOWEL APPLIANCES (Code for last 3 days)	*CODE:* 0. No 1. Yes a. Bedpan b. Enema c. Medication for control d. Ostomy	
6.	BOWEL APPLIANCE SUPPORT (Code for last 24 hours)	0. No appliances **(in item F5)** 1. Use of appliances, did not require help or supervision 2. Use of appliances, required supervision or setup 3. Minimal contact assistance (light touch only) 4. Moderate assistance; patient able to do 50% or more of tasks 5. Maximal assistance; patient able to do 25-49% of all sub-tasks 6. Total dependence	

Figure 3.3—MDS-PAC Section F: Bladder/Bowel Management

Table 3.1

Comparison of Scale and Item Means Using the Morris Translation

	FIM	MDS-PAC
Motor scale	45.46	50.26
Eating	5.51	5.96
Grooming	4.73	5.28
Bathing	3.24	3.39
Dressing upper body	4.25	4.66
Dressing lower body	2.99	3.28
Toileting	3.37	3.87
Bladder management	4.29	4.33
Bowel management	4.70	4.45
Transfer bed/chair	3.58	3.79
Transfer toilet	3.28	3.77
Transfer tub/shower	1.96	2.01
Locomotion—walk/wheelchair	2.22	3.78
Stairs	1.24	1.62
Cognitive scale	28.50	28.51
Comprehension	5.87	5.93
Expression	5.97	5.99
Social interaction	5.91	5.63
Problem solving	5.32	5.34
Memory	5.37	5.56

admission. Thus, this translation should be reviewed again before it is used for time points other than admission.

Admission Translation Rationale

The MDS-PAC Section E: Functional Status 3-Day ADL Self-Performance items employ an eight-point scale with ascending scores signifying greater dependence (0 = independent, 1 = setup help only, 2 = supervision, 3 = minimal assistance, 4 = moderate assistance, 5 = maximal assistance, 6 = total assistance, and 8 = activity did not occur). The FIM motor items use a seven-point scale with descending scores signifying greater dependence (7 = complete independence, 6 = modified independence, 5 = supervision or setup, 4 = minimal assistance, 3 = moderate assistance, 2 = maximal assistance, 1 = total assistance or activity did not occur). Thus, a first step in the translation is to rescore the MDS-PAC 0-6 to the approximate FIM scores (PAC 0 maps to FIM 7, 1 maps to 6, 2 to 5, 3 to 4, 4 to 3, 5 to 2, and 6 to 1). Further, it is generally agreed that the response category 8, did not occur, is used when individuals were unable to perform a task, so the MDS-PAC 8 is rescored to a FIM 1.

However, the translation between the two measures is much more complex than simply transposing these scores. To develop a meaningful translation between MDS-PAC and the FIM, consideration was given to several fundamental scoring differences between the measures including:

- Use of ADL Assist Codes,
- Scoring independence, and
- Scoring setup and supervision.

For instance, the FIM has one item for each motor activity, whereas the MDS-PAC has at least two items (3-Day ADL Self-Performance and ADL Assist Codes), and sometimes two or three additional items, that define a motor item. Thus, the translation between the FIM and MDS-PAC requires that, at a minimum, criteria include consideration of both the 3-Day ADL Self-Performance items and the ADL Assist Codes items. For some motor items, such as eating, transfers bed/chair, locomotion, bowel and bladder management, several additional MDS-PAC items must also be entered into the formula to convert MDS-PAC scores to FIM scores.

Use of ADL Assist Codes

The MDS-PAC 3-Day ADL Self-Performance items (E1) allow up to two exceptions where greater assistance was needed over the three-day assessment period, whereas the FIM is always graded for the most amount of assistance needed in a 24-hour period. To assist in converting MDS-PAC scores to FIM scores, a second set of MDS-PAC items, ADL Assist Codes (E2), provide additional information regarding one limb assistance (1) and two-person assistance (2) and are scored for the most dependent episode during the past three days. The ADL Assist Code section, particularly the one limb assist code, does not correspond precisely to the FIM assistance concepts and is therefore much more difficult to incorporate into the translation. Since one limb assist is weight-bearing assistance, the translation adopted the rule that the maximum FIM score a patient could have with an ADL Assist Code of 1 is a FIM 4, minimum assistance. Thus if the E1 item is scored 0, 1, or 2, which would convert to FIM 7, 6, or 5, but the ADL Assist Code is 1, then the item is rescored to FIM 4. For more dependent scores, MDS-PAC 3-6, an ADL Assist Code of 1 does not affect the scoring.

Because the FIM is a burden of care instrument, any item needing the assistance of two persons is always scored 1 (total assistance). The MDS-PAC specifies that two-helper assist must be used three or more times for maximal assistance and that total assistance (total dependence) requires full staff performance of activity during the entire period. Therefore, when a MDS-PAC ADL Assist Code for the item is scored 2 (2+ person assist), the translation automatically converts this to FIM score 1, regardless of the amount of subtasks that the patient performed.

Scoring Independence

The FIM differentiates between independence (grade 7) and modified independence (grade 6). A FIM grade 7 indicates that the activity was performed safely and independently and without assistive devices. Modified independence is used when there are safety concerns, or the patient requires extra time (three times normal), or the patient uses assistive devices to perform the activity independently. The MDS-PAC 3-Day Self-Performance motor items have a single grade for independence (grade 0), regardless of the equipment used or the manner in which the activity was performed. For some motor items, the MDS-PAC devices/aids items (E6) can be used to determine if an independent item should be scored as modified independence, a FIM score 6. For example, if the patient uses adaptive eating utensils and is scored independent, then the MDS-PAC score for eating is converted to a FIM score 6.

When the device item is not sufficient to separate cases that should be scored as modified independence from those that are truly independent, then both groups are scored at the level (FIM 6 or 7) that has been most frequently observed in the historical UDSmr FIM database at admission. Thus, if in the historical FIM database at admission more cases have been scored as modified independence (6) than as complete independence (7), the translation scores the group as a 6.

Scoring Setup and Supervision

The FIM has only a single grade that includes both set-up and supervision (5). The MDS-PAC has two grades that differentiate setup (1) from supervision (2). Thus, in the translation MDS-PAC scores of 1 and 2 are mapped to a FIM score 5.

Item-by-Item Translation

Tables 3.2 through 3.14 summarize the translation scoring rules for each motor item. Here, we provide the narrative to accompany and clarify the scoring rules in those tables.

Grooming, Bathing, Dressing Upper Body, Dressing Lower Body, Toileting, Transfer Toilet , and Transfer Tub/Shower

These motor items contain relatively simple translations, utilizing only two MDS-PAC items (3-Day ADL Self-Performance and ADL Assist Codes) to convert to FIM scores. The following general translation guidelines apply to the above seven motor items:

- If an item has the score of 2 in ADL Assist Code, then the converted FIM score is automatically total assistance (1) regardless of the score on the 3-Day ADL Self-Performance item.
- If both the 3-Day ADL Self-Performance and the ADL Assist Code for an item are scored 0, then the score is converted to a FIM score of 6 or 7 (each item uses the score that occurs most commonly in FIM). In this case, grooming and dressing upper body are converted to a FIM score 7. Bathing, dressing lower body, toileting, transfer toilet, and transfer tub/shower are converted to a FIM score 6.
- If an item is scored 1 or 2 on the 3-Day ADL Self-Performance and the ADL Assist Code is 0, then the converted FIM score is supervision or setup (5).
- If an item is scored 0, 1, or 2 on the 3-Day ADL Self-Performance and the ADL Assist Code is 1, then the converted FIM score is minimal assistance (4).
- If an item is scored 3 on the 3-Day ADL Self-Performance and the ADL Assist Code is 0 or 1, then the converted FIM score is minimal assistance (4).
- If an item is scored 4 on the 3-Day ADL Self-Performance and the ADL Assist Code is 1 or 0, then the converted FIM score is moderate assistance (3).
- If an item is scored 5 on the 3-Day ADL Self-Performance and the ADL Assist Code is 1 or 0, then the converted FIM score is maximal assistance (2).
- If an item is scored 6 or 8 on the 3-Day ADL Self-Performance regardless of how the ADL Assist Code is scored, then the converted FIM score is total assistance (1).
- If ADL Assist Code—grooming/personal hygiene (E2j) = 2, then FIM score = 1
- Else if
- If ADL Assist Code—bathing (E2k) = 2, then FIM score = 1
- Else if
- If ADL Assist Code—dressing upper body (E2e) = 2, then FIM score = 1
- Else if

Table 3.2

Grooming

3-Day ADL—Grooming Personal/Hygiene (E1j) =	ADL Assist Code—Grooming/Personal Hygiene (E2J) =	FIM Score
Independent (0)	Neither code (0)	Complete independence (7)
Independent (0)	Weight-bearing support 1 limb (1)	Minimal assistance (4)
Setup help only (1)	Neither code (0)	Supervision (5)
Setup help only (1)	Weight-bearing support 1 limb (1)	Minimal assistance (4)
Supervision (2)	Neither code (0)	Supervision (5)
Supervision (2)	Weight-bearing support 1 limb (1)	Minimal assistance (4)
Minimal assistance (3)	Neither code (0) or weight-bearing support 1 limb (1)	Minimal assistance (4)
Moderate assistance (4)	Neither code (0) or weight-bearing support 1 limb (1)	Moderate assistance (3)
Maximal assistance (5)	Neither code (0) or weight-bearing support 1 limb (1)	Maximal assistance (2)
Total assistance (6)	Neither code (0) or weight-bearing support 1 limb (1)	Total assistance (1)
Act did not occur (8)	Neither code (0) or weight-bearing support 1 limb (1)	Total assistance (1)

- If ADL Assist Code—dressing lower body (E2f) = 2, then FIM score = 1
- Else if
- If ADL Assist Code—toilet (E2h) = 2, then FIM score = 1
- Else if
- If ADL Assist Code—transfer toilet (E2i) = 2, then FIM score = 1
- Else if
- If ADL Assist Code—transfer tub/shower (E2l) = 2, then FIM score = 1
- Else if

Table 3.3

Bathing

3-Day ADL—Bathing (E1k) =	ADL Assist Code—Bathing (E2k) =	FIM Score
Independent (0)	Neither code (0)	Moderate independence (6)
Independent (0)	Weight-bearing support 1 limb (1)	Minimal assistance (4)
Setup help only (1)	Neither code (0)	Supervision (5)
Setup help only (1)	Weight-bearing support 1 limb (1)	Minimal assistance (4)
Supervision (2)	Neither code (0)	Supervision (5)
Supervision (2)	Weight-bearing support 1 limb (1)	Minimal assistance (4)
Minimal assistance (3)	Neither code (0) or weight-bearing support 1 limb (1)	Minimal assistance (4)
Moderate assistance (4)	Neither code (0) or weight-bearing support 1 limb (1)	Moderate assistance (3)
Maximal assistance (5)	Neither code (0) or weight-bearing support 1 limb (1)	Maximal assistance (2)
Total assistance (6)	Neither code (0) or weight-bearing support 1 limb (1)	Total assistance (1)
Act did not occur (8)	Neither code (0) or weight-bearing support 1 limb (1)	Total assistance (1)

Eating

According to FIM criteria, eating includes the use of suitable utensils to bring food to the mouth, chewing, and swallowing. To obtain a FIM score of 7 (total independence), the patient must be able to manage all food operations in a normal manner, safely, and independently.

In contrast, the MDS-PAC 3-Day ADL Self-Performance item for eating scoring allows a grade of independent (0) regardless of the skill or means with which a person eats. The translation, therefore, includes five MDS-PAC items that contribute information included in the single FIM item of eating: 3-Day ADL Self-Performance (E1g), ADL Assist Code (E2g), adaptive eating utensil (E6d), chewing problems (J1a), and swallowing (J2).

Table 3.4

Dressing Upper Body

3-Day ADL—Dressing Upper Body (E1e) =	ADL Assist Code—Dressing Upper Body (E2e) =	FIM Score
Independent (0)	Neither code (0)	Complete independence (7)
Independent (0)	Weight-bearing support 1 limb (1)	Minimal assistance (4)
Setup help only (1)	Neither code (0)	Supervision (5)
Setup help only (1)	Weight-bearing support 1 limb (1)	Minimal assistance (4)
Supervision (2)	Neither code (0)	Supervision (5)
Supervision (2)	Weight-bearing support 1 limb (1)	Minimal assistance (4)
Minimal assistance (3)	Neither code (0) or weight-bearing support 1 limb (1)	Minimal assistance (4)
Moderate assistance (4)	Neither code (0) or weight-bearing support 1 limb (1)	Moderate assistance (3)
Maximal assistance (5)	Neither code (0) or weight-bearing support 1 limb (1)	Maximal assistance (2)
Total assistance (6)	Neither code (0) or weight-bearing support 1 limb (1)	Total assistance (1)
Act did not occur (8)	Neither code (0) or weight-bearing support 1 limb (1)	Total assistance (1)

- If eating is scored 2 in ADL Assist Code, then the converted FIM score is automatically total assistance (1) regardless of the score on the 3-Day ADL Self-Performance eating item.
- If both the 3-Day ADL Self-Performance and the ADL Assist Codes for an item are scored 0, that item is converted to a FIM score 7. The exceptions to this rule are if adaptive eating utensils, or chewing problems, and/or swallowing problems are present, then the FIM score is 6.
- If eating is scored 1 or 2 on the 3-Day ADL Self-Performance and the ADL Assist Code is 0, then the converted FIM score is supervision or setup (5).
- If eating is scored 0, 1, or 2 on the 3-Day ADL Self-Performance and the ADL Assist Code is 1, then the converted FIM score is minimal assistance (4).
- If eating is scored 3 on the 3-Day ADL Self-Performance and the ADL Assist Code is 0 or 1, then the converted FIM score is minimal assistance (4).

Table 3.5

Dressing Lower Body

3-Day ADL—Dressing Lower Body (E1F) =	ADL Assist Code—Dressing Lower Body (E2f) =	FIM Score
Independent (0)	Neither code (0)	Moderate independence (6)
Independent (0)	Weight-bearing support 1 limb (1)	Minimal assistance (4)
Setup help only (1)	Neither code (0)	Supervision (5)
Setup help only (1)	Weight-bearing support 1 limb (1)	Minimal assistance (4)
Supervision (2)	Neither code (0)	Supervision (5)
Supervision (2)	Weight-bearing support 1 limb (1)	Minimal assistance (4)
Minimal assistance (3)	Neither code (0) or weight-bearing support 1 limb (1)	Minimal assistance (4)
Moderate assistance (4)	Neither code (0) or weight-bearing support 1 limb (1)	Moderate assistance (3)
Maximal assistance (5)	Neither code (0) or weight-bearing support 1 limb (1)	Maximal assistance (2)
Total assistance (6)	Neither code (0) or weight-bearing support 1 limb (1)	Total assistance (1)
Act did not occur (8)	Neither code (0) or weight-bearing support 1 limb (1)	Total assistance (1)

- If eating is scored 4 on the 3-Day ADL Self-Performance and the ADL Assist Code is 1 or 0, then the converted FIM score is moderate assistance (3).
- If eating is scored 5 on the 3-Day ADL Self-Performance and the ADL Assist Code is 1 or 0, then the converted FIM score is maximal assistance (2).
- If eating is scored 6 on the 3-Day ADL Self-Performance regardless of how the ADL Assist Code is scored, then the converted FIM score is total assistance (1).
- If ADL Assist Code—eating (E2g) = 2, then FIM score = 1
- Else if

Table 3.6

Toileting

3-Day ADL—Dressing Lower Body (E1h) =	ADL Assist Code—Dressing Lower Body (E2h) =	FIM Score
Independent (0)	Neither code (0)	Moderate independence (6)
Independent (0)	Weight-bearing support 1 limb (1)	Minimal assistance (4)
Setup help only (1)	Neither code (0)	Supervision (5)
Setup help only (1)	Weight-bearing support 1 limb (1)	Minimal assistance (4)
Supervision (2)	Neither code (0)	Supervision (5)
Supervision (2)	Weight-bearing support 1 limb (1)	Minimal assistance (4)
Minimal assistance (3)	Neither code (0) or weight-bearing support 1 limb (1)	Minimal assistance (4)
Moderate assistance (4)	Neither code (0) or weight-bearing support 1 limb (1)	Moderate assistance (3)
Maximal assistance (5)	Neither code (0) or weight-bearing support 1 limb (1)	Maximal assistance (2)
Total assistance (6)	Neither code (0) or weight-bearing support 1 limb (1)	Total assistance (1)
Act did not occur (8)	Neither code (0) or weight-bearing support 1 limb (1)	Total assistance (1)

Bladder Management

The FIM bladder management grade is based on both the control of the urinary bladder (frequency of incontinence) and level of assistance required for bladder management.

MDS-PAC contains three items that contribute information included in the single FIM item of bladder management: bladder continence (F1), bladder appliance (F2), and bladder appliance support (F3). This item includes both the episodes of incontinence as well as the assistance needed to manage urinary control.

Table 3.7

Transfer Toilet

3-Day ADL—Transfer Toilet (E1i)	ADL Assist Code—Transfer Toilet (E2i)	FIM Score
Independent (0)	Neither code (0)	Moderate independence (6)
Independent (0)	Weight-bearing support 1 limb (1)	Minimal assistance (4)
Setup help only (1)	Neither code (0)	Supervision (5)
Setup help only (1)	Weight-bearing support 1 limb (1)	Minimal assistance (4)
Supervision (2)	Neither code (0)	Supervision (5)
Supervision (2)	Weight-bearing support 1 limb (1)	Minimal assistance (4)
Minimal assistance (3)	Neither code (0) or weight-bearing support 1 limb (1)	Minimal assistance (4)
Moderate assistance (4)	Neither code (0) or weight-bearing support 1 limb (1)	Moderate assistance (3)
Maximal assistance (5)	Neither code (0) or weight-bearing support 1 limb (1)	Maximal assistance (2)
Total assistance (6)	Neither code (0) or weight-bearing support 1 limb (1)	Total assistance (1)
Act did not occur (8)	Neither code (0) or weight-bearing support 1 limb (1)	Total assistance (1)

- If both bladder continence and bladder appliance support are scored 0 and no bladder appliance is scored yes, then the converted FIM score is 7 (complete independence).
- If bladder continence is scored 0 and bladder appliance support is scored 1 and urinal/bedpan is no, then the converted FIM score is 6 (modified independence).
- If bladder continence is scored 0 and the only bladder appliance is medication and bladder appliance support is scored 0, 1, or 5, then the converted FIM score is 6 (modified independence). (MDS-PAC instructions score bladder appliance support a 5 when the nurse hands a patient a medication, whereas the FIM score would not be affected by the nurse handing patients medications.)

Table 3.8

Transfer Tub/Shower

3-Day ADL—Transfer Tub/Shower (E1l)	ADL Assist Code—Transfer Tub/Shower (E21)	FIM Score
Independent (0)	Neither code (0)	Moderate independence (6)
Independent (0)	Weight-bearing support 1 limb (1)	Minimal assistance (4)
Setup help only (1)	Neither code (0)	Supervision (5)
Setup help only (1)	Weight-bearing support 1 limb (1)	Minimal assistance (4)
Supervision (2)	Neither code (0)	Supervision (5)
Supervision (2)	Weight-bearing support 1 limb (1)	Minimal assistance (4)
Minimal assistance (3)	Neither code (0) or weight-bearing support 1 limb (1)	Minimal assistance (4)
Moderate assistance (4)	Neither code (0) or weight-bearing support 1 limb (1)	Moderate assistance (3)
Maximal assistance (5)	Neither code (0) or weight-bearing support 1 limb (1)	Maximal assistance (2)
Total assistance (6)	Neither code (0) or weight-bearing support 1 limb (1)	Total assistance (1)
Act did not occur (8)	Neither code (0) or weight-bearing support 1 limb (1)	Total assistance (1)

- If bladder continence is scored 0 and bladder appliance support is 0, and one or more of the following bladder appliances (external catheter, indwelling catheter, intermittent catheterization, ostomy, pads and/or briefs) is scored yes (1) and urinal/bedpan is coded no, then the converted FIM score is 6 (modified independence).
- If bladder continence is scored 1 and bladder appliance support is scored 0 or 1 and urinal/bedpan is coded no, then the converted FIM score is 6 (modified independence).
- If bladder continence is scored 0 or 1 and the only bladder appliance is urinal/bedpan and bladder appliance support is scored 0 or 1, then the converted FIM score is 5 (setup or supervision).

Table 3.9

Eating

3-Day ADL—Eating (E1g) =	ADL Assist Code—Eating (E2g) =	Devices/ Aids—Adaptive Eating Utensil (E6d) =	Oral Problems—Chewing Problems (J1a) =	Oral Problems—Swallowing (J2) =	FIM Score
Independent (0)	Neither code (0)	Unchecked	Normal	Normal (0)	Complete independence (7)
Independent (0)	Neither code (0)	Checked	Normal (0)	Normal (0)	Moderate independence (6)
Independent (0)	Neither code (0)	Unchecked	Yes (1) or	(1) or (2) or (3) or (4)	Moderate independence (6)
Independent (0)	Neither code (0)	Checked	Yes (1) or	(1) or (2) or (3) or (4)	Moderate independence (6)
Independent (0)	Weight-bearing support 1 limb (1)				Minimal assistance (4)
Setup help only (1)	Neither code (0)				Supervision (5)
Setup help only (1)	Weight-bearing support 1 limb (1)				Minimal assistance (4)
Supervision (2)	Neither code (0)				Supervision (5)
Supervision (2)	Weight-bearing support 1 limb (1)				Minimal assistance (4)
Minimal assistance (3)	Neither code (0) or weight-bearing support 1 limb				Minimal assistance (4)
Moderate assistance (4)	Neither code (0) or weight-bearing support 1 limb				Moderate assistance (3)
Maximal assistance (5)	Neither code (0) or weight-bearing support 1 limb				Maximal assistance (2)
Total assistance (6)	Neither code (0) or weight-bearing support 1 limb				Total assistance (1)
Act did not occur (8)	Neither code (0) or weight-bearing support 1 limb				Total assistance (1)

Table 3.10

Bladder Management

Maximum of Control of Urinary Bladder Function (F1a) and Bladder Appliance Support (F3)	Control of Urinary Bladder Function (F1a)	Maximum of Bladder Appliances (F2a, F2b, F2c, F2d, F2e, F2f)	Bladder Appliance: Medications for Control (F2d) Only	Bladder Appliance: Urinal, Bedpan (F2g)	Bladder Appliance Support (F3)	FIM Score
(F1a): Continent (0) (F3): No appliances (0)	Continent (0)	No (0)	No (0)	No (0)	No appliances (0)	Complete independence (7)
	Continent (0)	No (0)	Yes (1)	No (0)	Maximal assistance (5)	Moderate independence (6)
(F1a): Continent with catheter (1) (F3): Use of appliances, no help or support (1)				No (0)		Moderate independence (6)
	Continent (0)	Yes (1)		No (0)	No appliances (0) or use of appliance, no support (1)	Moderate independence (6)
	Continent (0) or continent with catheter (1)			Yes (1)	No appliances (0) or use of appliance, no support (1)	Supervision (5)
(F1a): Biweekly incontinence (2) (F3): Use of appliances, requires support (2)						Supervision (5)
(F1a): Weekly incontinence (3) (F3): Minimal assistance (3)						Minimal assistance (4)
(F1a): Occasionally incontinent (4) (F3): Moderate assistance (4)						Moderate assistance (3)
(F1a): Frequently incontinent (5) (F3): Maximal assistance (5)						Maximal assistance (2)
(F1a): Incontinent (6) (F3): Total dependence (6)						Total assistance (1)
(F1a) and (F3): Did not occur (8)						Total assistance (1)

- If bladder continence is scored 2 and bladder appliance support is 0, 1, or 2, then the converted FIM score is 5 (setup or supervision).
- If the largest score between bladder continence and bladder appliance support is 3, then the converted FIM score is 4 (minimal assistance).
- If the largest score between bladder continence and bladder appliance support is 4, then the converted FIM score is 3 (moderate assistance).
- If the largest score between bladder continence and bladder appliance support is 5, then the converted FIM score is 2 (maximal assistance).
- If the largest score between bladder continence and bladder appliance support is 6 or 8, then the converted FIM score is 1 (total assistance).

Bowel Management

MDS-PAC contains three items that contribute information included in the single FIM item of bowel management: bowel continence (F4), bowel appliance (F5), and bowel appliance support (F6).

- If bowel continence is scored 0 and bowel appliance support is scored 0 and no bowel appliances is scored yes, then the FIM score is 7 (independence).
- If bowel continence is scored 0 and bowel appliance support is scored 0 and ostomy F5d is yes, then the FIM score is 6 (modified independence).
- If bowel continence is scored 0 and bowel appliance support is scored 0, 1, 5, or 6 and medications is the only bowel appliance scored yes, then the FIM score is 6 (modified independence).
- If bowel continence is scored 0 and bowel appliance support is scored 1 and neither bedpan nor enema is yes, then the FIM score is 6 (modified independence).
- If bowel continence is scored 1 and bowel appliance support is scored 0 or 1 and neither bedpan nor enema is yes, then the FIM score is 6 (modified independence).
- If bowel continence is scored 0 or 1 and bowel appliance support is scored 0 or 1 and bedpan is checked, then the FIM score is 5 (setup or supervision).
- If the largest score between bowel continence and bowel appliance support is 2 and enema is no, then the FIM score is 5 (setup or supervision).
- If enema is indicated and the largest score between bowel continence and bowel appliance support is 3, then the FIM score is 4 (minimal assistance).

Table 3.11

Bowel Management

Maximum of Bowel Continence (F4) and Bowel Appliance Support (F6)	Bowel Continence (F4)	Bowel Appliances Bedpan (F5a)	Bowel Appliances Enema (F5b)	Bowel Appliance: Medications for Control (F5c)	Bowel Appliance: Ostomy (F5d)	Bowel Appliance Support (F6)	FIM Score
(F4): Continent (0) (F6): No appliances (0)		No (0)	No (0)	No (0)	No (0)		Complete independence (7)
(F4): Continent with ostomy (1) (F6): Use of appliances, no help or support (1)		No (0)	No (0)				Moderate independence (6)
	Continent (0)	No (0)	No (0)	Yes (1)	No (0)		Moderate independence (6)
(F4): Biweekly incontinence (2) (F6): Use of appliances, requires help or support (2)			No (0)				Supervision (5)
	Continent (0) or continent with ostomy (1)	Yes (1)	No (0)	No (0) or Yes (1)	No (0) or (1)	(0) or (1) or (2)	Supervision (5)
	Continent (0) or continent with ostomy (1) or biweekly incontinence (2)	No (0) or Yes (1)	Yes (1)	No (0) or Yes (1)	No (0) or (1)	(0) or (1) or (2)	Minimal assistance (4)
(F4): Weekly incontinence (3) (F6): Minimal contact assistance (3)							Minimal assistance (4)
(F4): Occasionally incontinent (4) (F6): Moderate assistance (4)							Moderate assistance (3)
(F4): Frequently incontinent (5) (F6): Maximal assistance (5)							Maximal assistance (2)
(F4): Incontinent (6) (F6): Total dependence (6)							Total assistance (1)
(F4) Did not occur (8)							Total assistance (1)

- If the largest score between bowel continence and bowel appliance support is 3, then the FIM score is 4 (minimal assistance).
- If the largest score between bowel continence and bowel appliance support is 4, then the FIM score is 3 (moderate assistance).
- If the largest score between bowel continence and bowel appliance support is 5, then the FIM score is 2 (maximal assistance).
- If the largest score between bowel continence and bowel appliance support is 6 or 8, then the FIM score is 1 (total assistance).

Transfer Bed, Chair, Wheelchair

The scoring criteria for transfers is the same as for the first eight items (grooming, etc.) with the addition of devices/aids (cane, crutch, walker, or orthotics/ prosthetics) as assistive devices. When any of these are identified and both the 3-Day ADL Self-Performance items and the ADL Assist Codes for an item are scored 0, then item is converted to a FIM score of 6. The rationale is that these devices are typically used for transfers.

- If ADL Assist Code—transfer bed/chair (E2b) = 2, then FIM score = 1
- If devices/aids—mechanical lift (E6e) is checked, then FIM score = 1
- Else if

Locomotion

The FIM locomotion item is based on the patient's usual mode of locomotion (walk or wheelchair) that is anticipated at discharge and uses distance as part of the scoring metric. The FIM requires that the patient travel at least 150 feet to be graded higher than a 2 (maximal assistance), with the exception of household ambulation. The MDS-PAC 3-Day ADL Self-Performance locomotion item does not differentiate walk from wheelchair and does not contain any information about distance. However, walking distance is scored in the MDS-PAC item distance walked (E8a). Because 87 percent of the FIM scores for locomotion identified walking as the mode of locomotion and because there is no way to identify distance traveled in a wheelchair, the translation utilizes the following four items in the locomotion translation: MDS-PAC 3-Day ADL Self-Performance item walk in facility (E1d), ADL Assist Code for walk in facility (E2d), distance walked (E8a), and the devices/aids item (E6a, b, or f).

Table 3.12

Transfer Bed, Chair, Wheelchair

3-Day ADL—Transfer Bed/Chair (E1b)	ADL Assist Code—Transfer Bed/Chair (E2b)	Devices/Aids—Cane/Crutch (E6a) =	Devices/Aids—Walker (E6b) =	Devices/Aids—Slide Board (E6h)=	FIM Score
Independent (0)	Neither code (0)	Not checked	Not checked	Not checked	Complete independence (7)
Independent (0)	Neither code (0)	Checked or	Checked or	Checked	Moderate independence (6)
Independent (0)	Weight-bearing support 1 limb (1)				Minimal assistance (4)
Setup help only (1)	Neither code (0)				Supervision (5)
Setup help only (1)	Weight-bearing support 1 limb (1)				Minimal assistance (4)
Supervision (2)	Neither code (0)				Supervision (5)
Supervision (2)	Weight-bearing support 1 limb (1)				Minimal assistance (4)
Minimal assistance (3)	Neither code (0) or weight-bearing support 1 limb (1)				Minimal assistance (4)
Moderate assistance (4)	Neither code (0) or weight-bearing support 1 limb (1)				Moderate assistance (3)
Maximal assistance (5)	Neither code (0) or weight-bearing support 1 limb (1)				Maximal assistance (2)
Total assistance (6)	Neither code (0) or weight-bearing support 1 limb (1)				Total assistance (1)
Activity did not occur (8)	Neither code (0) or weight-bearing support 1 limb (1)				Total assistance (1)

Table 3.13

Locomotion—Walk/Wheelchair

3-Day ADL—Walk in Facility (E1d)	ADL Assist Code—Walk in Facility (E2d)	Farthest Distance Walked (feet) (E8a)	Devices/Aids (E6)	FIM Score
Independent (0)	Neither code (0)	150+ (0)	Unchecked: cane/crutch (E6a) and walker (E6b) and orthotics/ prosthesis (E6f)	Complete independence (7)
Independent (0)	Neither code (0)	150+ (0)	Checked: cane/crutch (E6a) or walker (E6b) or orthotics/ prosthesis (E6f)	Moderate independence (6)
Independent (0)	Neither code (0)	51–149 (1)		Household ambulation exception (5)
Independent (0)	Weight-bearing support 1 limb (1)	150+ (0)		Minimal assistance (4)
Setup help only (1)	Neither code (0)	150+ (0)		Supervision (5)
Setup help only (1)	Weight-bearing support 1 limb (1)	150+ (0)		Minimal assistance (4)
Supervision (2)	Neither code (0)	150+ (0)		Supervision (5)
Supervision (2)	Weight-bearing support 1 limb (1)	150+ (0)		Minimal assistance (4)
Supervision (2)	Neither code (0) or weight-bearing support 1 limb (1)	51–149 (1)		Maximal assistance (2)
Minimal assistance (3)	Neither code (0)	150+ (0)		Minimal assistance (4)

Table 3.13 (continued)

3-Day ADL—Walk in Facility (E1d)	ADL Assist Code—Walk in Facility (E2d)	Farthest Distance Walked (feet) (E8a)	Devices/Aids (E6)	FIM Score
Minimal assistance (3)	Weight-bearing support 1 limb (1)	150+ (0)		Minimal assistance (4)
Minimal assistance (3)	Neither code (0) or weight-bearing support 1 limb (1)	51–149 (1)		Maximal assistance (2)
Moderate assistance (4)	Neither code (0) or weight-bearing support 1 limb (1)	150+ (0)		Moderate assistance (3)
Moderate assistance (4)	Neither code (0) or weight-bearing support 1 limb (1)	51–149 (1)		Maximal assistance (2)
Maximal assistance (5)	Neither code (0) or weight-bearing support 1 limb (1)	150+ (0) or 51–149 (1)		Maximal assistance (2)
Total assistance (6)	Neither code (0) or weight-bearing support 1 limb (1)	150+ (0) or 51–149 (1)		Total assistance (1)
Act did not occur (8)	Neither code (0) or weight-bearing support 1 limb (1)	150+ (0) or 51–149 (1)		Total assistance (1)

- If the MDS-PAC 3-Day ADL Self-Performance item (E2d) is scored a 2, then the FIM score is 1 (total assistance).
- If the distance walked (E8a) is greater than 1, then the FIM score is 1 (total assistance).
- If both the MDS-PAC 3-Day ADL Self-Performance item walk in facility (E1d) and the ADL Assist Code (E2d) are scored 0, and distance walked (E8a) is scored 0 (> 150 feet), and the devices/aids items E6a, b, and f are not checked, then the converted FIM score is 7 (complete independence).
- If both the MDS-PAC 3-Day ADL Self-Performance item walk in facility (E1d) and the ADL Assist Code (E2d) are scored 0, and distance walked (E8a) is scored 0 (> 150 feet), and the devices/aids items E6a, b, or f are checked, then the converted FIM score is 6 (modified independence).
- If both the MDS-PAC 3-Day ADL Self-Performance item walk in facility (E1d) and the ADL Assist Code (E2d) are scored 0, and distance walked (E8a) is scored 1 (50 to 149 feet), then the converted FIM score is 5 (household ambulation exception).
- If the MDS-PAC 3-Day ADL Self-Performance item walk in facility (E1d) is scored 1 or 2, and the ADL Assist Code (E2d) is scored 0, and distance walked (E8a) is scored 0 (> 150 feet), then the converted FIM score is 5 (supervision or setup).
- If the MDS-PAC 3-Day ADL Self-Performance item walk in facility (E1d) is scored 0, 1, or 2, and the ADL Assist Code (E2d) is scored 1, and distance walked (E8a) is scored 0 (> 150 feet), then the converted FIM score is 4 (minimal assistance).
- If the MDS-PAC 3-Day ADL Self-Performance item walk in facility (E1d) is scored 3, and the ADL Assist Code (E2d) is scored 0 or 1, and distance walked (E8a) is scored 0 (> 150 feet), then the converted FIM score is 4 (minimal assistance).
- If the MDS-PAC 3-Day ADL Self-Performance item walk in facility (E1d) is scored 4, and the ADL Assist Code (E2d) is scored 0 or 1, and distance walked (E8a) is scored 0 (> 150 feet), then the converted FIM score is 3 (moderate assistance).
- If the MDS-PAC 3-Day ADL Self-Performance item walk in facility (E1d) is scored 5, and the ADL Assist Code (E2d) is scored 0 or 1, and distance walked (E8a) is scored 0 (> 150 feet), then the converted FIM score is 2 (maximal assistance).
- If the MDS-PAC 3-Day ADL Self-Performance item walk in facility (E1d) is scored 1, 2, 3, 4, or 5, and the ADL Assist Code (E2d) is scored 0 or 1, and

distance walked (E8a) is scored 1 (50 to 149 feet), then the converted FIM score is 2 (maximal assistance).

- If the MDS-PAC 3-Day ADL Self-Performance item walk in facility (E1d) is scored 6 or 8, regardless of the scores of either ADL Assist Code (E2d) or distance walked (E8a), then the converted FIM score is 1 (total assistance).
- If ADL Assist Code—walk in facility (E2d) = 2, then FIM score = 1
- Else if farthest distance walked without sitting down (E8a) > 1, then FIM score = 1
- Else if

Stairs

Stairs is the least complex of all items because the MDS-PAC stair-climbing item codes for the most dependent episode in the last 24 hours. After converting any MDS-PAC item coded 8 to a FIM score 1, all other MDS-PAC scores are simply inverted to convert to FIM scores (0 = 7, 1 = 6, 2 = 5, 3 = 4, 4 = 3, 5 = 2, 6 = 1).

Table 3.14

Stairs

Walking and Stair Climbing—Stair Climbing (E8c)	FIM Score
Complete independence (0)	Complete independence (7)
Modified independence (1)	Modified independence (6)
Supervision (2)	Supervision (5)
Minimal assistance (3)	Minimal assistance (4)
Moderate assistance (4)	Moderate assistance (3)
Maximal assistance (5)	Maximal assistance (2)
Total assistance (6)	Total assistance (1)
Activity did not occur in last 24 hours (8)	Total assistance (1)

Scoring Differences That Could Not Be Corrected in the Translation

General Limitations

Assessment Period. One fundamental difference between the MDS-PAC and the FIM is the assessment period. The FIM is a performance measure in which the clinician scores the patient's most dependent episode for each activity during a 24-hour period. The FIM must be completed within 72 hours of admission, thus the exact 24-hour period used to grade each item is unclear. In contrast, the MDS-PAC uses a three-day look-back period that requires knowledge of the amount of support needed for all episodes of each activity over all shifts throughout the three-day assessment period. The rater must consider the three most dependent episodes occurring during the three days. Information is gathered from the patient and family, direct-care providers, and medical record as well as the rater's own observations. The number of times the activity occurred and the amount of assistance needed during each episode must be considered in calculating a grade for the MDS-PAC.

Because MDS-PAC scoring of the 3-Day ADL Self-Performance items considers all episodes of the activity throughout three days, the scoring provides exceptions for occurrences of heavier care. For example, if an item is coded 0 (independent), 1 (setup help only), 2 (supervision), or 3 (minimal assistance), then there can have been only one or two occurrences where weight-bearing support was needed. However, if heavier care was needed three or more times, the item is scored 4 (moderate assistance) or 5 (maximal assistance), for the amount of assistance that occurred most often. If performance of an item fluctuated during the three-day period and no one type of assistance was provided to the patient three or more times, then the code for the least dependent episode applies. The two exceptions to this rule are the bathing and transfer tub/shower items, which are graded for the episode in which the heaviest amount of assistance was needed.

ADL Assist Codes. The ADL Assist Codes are of limited value in determining the amount of assistance needed. When the ADL Assist Code of 1 (one limb assistance) is used, it is impossible to know whether minimal, moderate, or maximal assistance was required. In addition, ADL Assist code 0 is used in a majority of situations. A code of 0 may signify that no assist was needed, or that two limb or trunk assistance was needed. In fact, since one limb assistance is a relatively infrequent type of assistance, the 0 code for this item is used in most cases, unless two or more persons assist. Therefore, the ADL Assist Codes are insufficient to precisely score the amount of assistance needed for the most dependent episode. The most frequently used code, 0, is not helpful at all in the translation. It is very likely, in fact, that many items that would have been converted to a lower score may be missed because of the insufficient scoring criteria used for ADL Assist Codes.

Scoring Independence. MDS-PAC is unable to differentiate normal motor tasks from poor skill or the need for excessive time to complete the task. Therefore, it is quite possible that many items with the FIM score of modified independence (grade 6) may be scored as total independence on the MDS-PAC (grade 0) and therefore converted to a FIM score 7. In addition, the translation was unable to differentiate between FIM scores 6 and 7 for some items and therefore only six of the seven grades on the FIM scale can be utilized. Bathing, toileting, dressing lower body, transfer toilet, transfer tub/shower cannot have a FIM score 7 (complete independence). Likewise, two items—grooming and dressing upper body—cannot be converted to FIM score 6 (modified independence).

Scoring Total Assistance. A motor item receives a FIM score 1 (total assistance) when the patient provides less than 25 percent of the tasks. The MDS-PAC is coded a 6 (total assistance) only when the patient does not participate in the activity at all and requires full staff performance. This discrepancy in scoring criteria could result in higher MDS-PAC scores.

In addition, the MDS-PAC ADL Assist Codes are useful in identifying whether two-person assistance was utilized during the most dependent episode within the past three days. However, it is not possible to identify that the most amount of assistance scored in this section occurred within the same 24-hour period as the FIM assessment period.

Limitations in Converting MDS-PAC to FIM Scores for Specific Items

Locomotion: Walk/Wheelchair. According to the MDS-PAC definition, to receive a score other than an 8 (activity did not occur) for the walk in facility item, the patient must walk in places other than within their room only. The FIM guidelines do not specify the location where walking is performed. Thus, it is possible for some MDS-PAC items to be scored 8, when the FIM may have been scored differently. However, since FIM has a 150-foot criterion, this may be an insignificant distinction, except for the grades that use 50 feet as the criterion. We explored using the MDS-PAC item locomotion for the translation, since this MDS-PAC item includes use of a wheelchair or walking, but it was not as effective as the walk in facility item. However, by using just the walk in facility item in the translation, we cannot differentiate wheelchair users from ambulators in the translation.

Bladder and Bowel Management. Unfortunately, because of the manner in which the items were organized, it is sometimes impossible to distinguish which bladder or bowel appliance is being graded for support. For instance, if a patient used a bedpan and also was receiving medication for control, the bladder appliance support item would reflect only the amount of assistance for the task that occurred most frequently. Therefore, if multiple bladder (or bowel) appliances were coded yes, there is no way of knowing which appliance was being graded in the bladder (or bowel) appliance support item.

Another problem for this item is that grading for medication in the bladder (or bowel) appliance support section requires a grade of maximal assistance to be assigned when the nurse passes the medication to the patient. Medications are routinely controlled by nursing staff in the acute rehabilitation setting, regardless of the patient's ability to participate in this activity.

Transfer Toilet. The MDS-PAC includes transfers on and off the bedpan as part of the transfer toilet item. A patient who needs physical assistance to get on and off the bedpan will receive a MDS-PAC grade that reflects a physical assistance for the bedpan transfer. Transferring on and off the bedpan is not considered part of transfer toilet activity in FIM scoring.

Additional Comments

1. The FIM definition of toileting and the MDS-PAC definition of toilet use are not exactly the same. The MDS-PAC definition includes the following tasks that are *not* included in the FIM definition: changes pad, manages ostomy or catheter.
2. The FIM and MDS-PAC interpret assistance from two or more persons differently. On the MDS-PAC, two-person assistance means that *both* helpers are providing hands-on assistance. For the FIM instrument, two-person assistance would include one helper providing hands-on assistance while a second helper provides standby assistance.
3. There are different rules for scoring for the 3-Day ADL Self-Performance items, which are based on the level of assistance that occurs *three or more times*, and the ADL Assist Codes, which are based on the *most dependent* level of assistance. In some cases, if the items are scored accurately, the patient may be independent most of the time (3-Day Self-Performance = 0), but need limb support once (ADL Assist Code = 1).
4. For bladder management, the time frames for continence and bladder appliance and bladder appliance support are different. Continence is 14 days and the appliance items reflect the last 24 hours only.
5. Bladder and bowel FIM items have been more difficult than other FIM items for clinical staff to accurately score. In the MDS-PAC, this problem appears to be inflated for bladder management, since the clinicians would need to

evaluate three items with multiple possible scoring combinations. Essentially, the tool seems overly complicated for assessing this domain.

6. If the bowel medication is a suppository, then FIM scores could range from 6 (patient administers him/herself) to level 4 (helper administers medication; patient requires no further assistance with the bowel program.

Summary

We began with a set of translation rules (the Morris translation) that took MDS-PAC items and converted them into the 18 FIM items. We found that mean cognitive scale scores between the FIM and its MDS-PAC translation were quite comparable but that the mean motor scale scores were not. After reviewing the scoring rules of both instruments, we developed a new admission translation for the motor items but retained the original translation for the cognitive items. The following section describes the new scoring rules and identifies differences between the two instruments that could not be overcome with our scoring rules.

4. Conclusions

This section presents some basic characteristics of the two instruments. Our focus is often limited to the items that will be used in the new payment system. The first and perhaps most fundamental piece of information needed is the reason for the admission to inpatient rehabilitation, so we begin by looking at the completeness and quality of these data. Information on form completeness and data entry for other payment items follows. Analyses then compare inter-team rescoring reliability on each instrument. Factor analysis is used to compare whether the MDS-PAC and the FIM are measuring the same concepts and to evaluate how the translation affects the concepts being measured. Last, the administrative burden (completion times) of the two instruments is compared.

Coding the Reason for a Rehabilitation Admission

In short term acute hospitals, diagnoses are used to describe a patient's condition and the primary diagnosis describes the reason for admission. Diagnoses have been standardized and are often communicated using a system called the International Classification of Diseases, Version 9 (ICD-9). In rehabilitation hospitals, impairment groups serve this function. The UDSmr system has 17 broad impairment groupings, such as stroke, brain dysfunction, spinal cord dysfunction, amputation of limb, arthritis, cardiac disorders, and medically complex cases. Each grouping has a number of impairment groups within it. For example, under stroke there are five impairment groups, left body (right brain) stroke, right body (left brain) stroke, bilateral stroke, stroke with no paresis, and other stroke. These describe the affected parts of the body that are important for rehabilitation. Etiologic diagnoses describe the cause of the impairment and may be associated either with the broad grouping (stroke) or with the actual impairment groups (left body (right brain) stroke, right body (left brain) stroke, etc.). For stroke, etiologic diagnoses include subarachnoid hemorrhage, intracerebral hemorrhage, intracranial hemorrhage, occlusion or stenosis of precerebral arteries, etc., which describe the location of the hemorrhage or occlusion and can be the cause of any of the stroke impairment groups as the latter describe the affected body parts. Etiologic diagnoses are used to help therapists identify the appropriate broad grouping or actual impairment group. Each impairment group is associated with a rehabilitation impairment category (RIC) and 21 RICs form the basic structure or first tier of the CMGs in the payment system. In many cases, such as stroke, the RIC and the broad impairment grouping contain the same set of impairment groups. However, in a few cases the RIC breaks out or reassigns impairment groups from their broad impairment grouping. For example, Guillain-Barre is in the broad impairment grouping called neurological conditions but forms its own RIC.

When therapists are trained on the FIM by UDSmr, they are also instructed in how to code impairment groups and are given a coding manual that lists the impairment groups and the associated etiologic diagnoses. The form where the FIM is recorded also includes an item for impairment group and another for the etiologic diagnosis. An impairment group coding manual is also included in the MDS-PAC manual. The MDS-PAC has an item for impairment group but does not ask for an etiologic diagnosis. It does include both a disease checklist and an item asking for "Other current or more detailed diagnoses and ICD-9 codes." When therapists and nurses were trained on the MDS-PAC, they were told that the impairment group item was from the FIM system.

For this study, both FIM and MDS-PAC data were collected in handwritten format on study forms and later entered into a machine- readable format. Although data entry software has not been developed for the MDS-PAC, UDSmr routinely collects FIM data through FIMware. Unfortunately, this software was not a realistic option for this study.

Each MDS-PAC, MDS-PAC Face Sheet, FIM and FIM Face Sheet study form was pre-labeled with a patient identification number for the study. Since the forms were separated during data collection, we recorded a small set of patient characteristics (birth date, admission date, zip code of residence, and gender) on the face sheets so that we could be sure that the re-linked data actually referred to the same patient. For each set of matched labels, we compared the set of patient characteristics. When these failed to match, we reviewed the forms and mismatched data. Where discrepancies appeared to be simple transpositions of digits, or common date substitutions, we allowed the match. If one form was missing or if we could not reconcile the patient characteristics on matching face sheets, these cases were not included in our matched samples. This reduced the available number of cases by five.

An important feature of FIMware is that it provides a menu of rehabilitation impairment groups, which allows for easy entry, rather than relying on a therapist's written report for the code. This feature ensures that each form contains a valid impairment group, although it does not ensure that the selected code is accurate. (One step toward verifying accuracy would be to check that the etiologic diagnosis actually supports the selected impairment code.) Because the study was not able to take advantage of this software, we encountered some missing data problems typically not experienced by UDSmr. We found that the impairment group field contained both blanks and invalid codes. Some of these invalid codes were the result of data entry errors where the problem was merely a misplaced decimal or a lack of specificity that would not change the RIC assignment. When an impairment code was missing or invalid and the etiologic diagnosis clearly pointed to single RIC assignment, we corrected these. In addition, data collectors were sometimes confused by the distinction between impairment codes and RICs. Some supplied both on the form and when we were able to sort the two numbers out, these forms were corrected as well. Clearly, our data quality would have been better had we supplied a list of valid impairment codes on each form.

In Table 4.1, we report the percentage of cases with missing/invalid values for the impairment group fields. Here, we found that for both forms, 10–11 percent of cases had missing or invalid impairment code data. For many of the study analyses, we needed data on both the FIM and the MDS-PAC, so forms that were missing the impairment group on one or the other were unusable and our initial matched sample loss was 15 percent. After correcting these codes, we found that 91 percent of the institutional sample with both forms had impairment groups that mapped to the same RIC.

We had not anticipated these difficulties in collecting impairment group information. To explore further, we include data from the calibration teams and compare the levels of disagreement on RIC selection for different combinations of scoring teams and instruments. These are shown in Table 4.2. We found the best agreement (95 percent) when both the FIM and the MDS-PAC were scored by the

Table 4.1

Number and Percentage of Cases with Usable Impairment Code Data

	FIM	PAC	FIM and MDS-PAC Match
Total number of cases	3,540	3,497	3,484
Cases with missing/invalid impairment codes	363 (10%)	401 (11%)	520 (15%)
Corrected impairment codes	197 (6%)	232 (7%)	317 (9%)
Cases with valid impairment codes after cleaning	3,374 (95%)	3,328 (95%)	3,281 (94%)
Cases with matched RICs			3,181 (91%)

Table 4.2

Disagreement on RIC Selection by Instrument and Scoring Team Combination

	No. of Cases	Percentage Disagreement
Same institutional team, FIM vs. PAC	2,372	8
Different institutional teams, FIM vs. PAC	1,056	9
Unknown institutional teams, FIM vs. PAC	52	17
Calibration teams, FIM vs. PAC	241	5
Institutional vs. calibration team, FIM	239	29
Institutional vs. calibration team, PAC	239	27

calibration teams. When the institutional teams scored the FIM and the MDS-PAC, we expected to see more disagreement when the scoring of the two instruments was done by different teams. However, our data showed only a small difference, 8 vs. 9 percent. In the small percentage of cases where the forms were missing data on whether the same or different institutional teams scored the two instruments, disagreement was higher, 17 percent. The highest levels of disagreement occurred when we compared the RICs for the calibration team cases with those for the institutional cases. Here, 27–29 percent of the cases disagreed. This suggests that there is a substantial amount of ambiguity in impairment group assignment and that clearer instructions for assignment are warranted.

To see if some RICs seemed to have more scoring error than others, we treated the FIM RIC as the base and calculated the percentage of PAC cases with the same RIC when the institutional teams scored the two instruments (~ 3,300 cases) and for one set of comparisons between the institutional teams and the calibration teams (230 cases). These are shown in Table 4.3. Agreement is lowest for RICs 13, rheumatoid and other arthritis, and 18, multiple major trauma with brain and/or spinal cord injury (74 and 67 percent, respectively), both small RICs that are not even represented in the calibration team sample. Of greater concern are the relatively low levels of agreement between the calibration teams and the institutional teams for RICs in which the calibration team had a reasonable number of cases. This would include RICs 7, 8, 9, and 20 and perhaps even RICs 14 and 15.

Table 4.3

RIC Agreement Between Institutional and Calibration Teams

RIC	Percentage Agreement for Institutional Teams, FIM vs. MDS-PAC	Percentage Agreement for Institutional Team vs. Calibration Team
1 Stroke	97	100
2 Traumatic brain injury	95	—
3 Non-traumatic brain injury	92	80
4 Traumatic spinal cord injury	83	50
5 Non-traumatic spinal cord injury	88	86
6 Neurological	97	75
7 Lower extremity fracture	94	82
8 Lower extremity joint replacement	96	86
9 Other orthopedic	92	30
10 Amputation, lower extremity	98	100
11 Amputation, other	88	—
12 Osteoarthritis	90	50
13 Rheumatoid, other arthritis	74	—
14 Cardiac	94	57
15 Pulmonary	89	70
16 Pain syndrome	86	0
17 MMT[a]—NBSCI[b]	93	100
18 MMT—BSCI[c]	67	—
19 Guillain-Barre	100	—
20 Miscellaneous	94	60

[a]MMT = multiple major trauma.
[b]NBSCI = no brain or spinal cord injury.
[c]BSCI = brain or spinal cord injury.

Completeness of Other Items

For functional status items, the percentage of data missing is generally low but notably higher for the bowel and bladder items, especially on the FIM. These are both more complicated items to score and are most frequently scored by nursing rather than rehabilitation therapists. Although the therapists tend to be located together, this is often not on the nursing unit, so getting the form to nursing and back presented an extra step, which was sometimes forgotten. MDS-PAC data were always collected by study-trained teams, but the FIM data were obtained from all FIM-certified and -licensed clinicians in the facility. Some of these individuals were not as vested in the study as our primary data collectors. Although the percentage missing on any one item is low, the cumulative effect is somewhat larger resulting in a loss of 5.0 to 5.9 percent of the cases (see Table 4.4). When we consider only matched cases, that is, those with both a FIM and MDS-PAC for the same patient and a valid RIC, missing motor and cognitive scores reduced the sample 10 percent more, from 3,281 to 2,953.

Table 4.4

Number and Percentage of Cases with Missing Functional and Cognitive Status Data

	FIM Percentage Missing	PAC Percentage Missing	Matched FIM and MDS-PAC Percentage Missing
Total number of cases	3,540	3,498	3,484
Number of cases with valid RICs	3,374	3,328	3,281
Number of cases with valid matched RICs			3,181
% motor score missing	3.8	3.5	
% cognitive score missing	2.1	2.7	
% motor or cognitive missing	5.0	5.9	
Total cases with motor and cognitive scores	3,324	3,287	
% eating missing	0.7	0.4	
% grooming missing	0.7	0.4	
% bathing missing	0.7	0.5	
% dressing upper body missing	0.7	0.3	
% dressing lower body missing	0.7	0.4	
% toileting missing	1.1	0.6	
% bladder management missing	2.2	0.8	
% bowel management missing	2.4	0.8	
% transfer bed/chair missing	0.7	0.5	
% transfer toilet missing	0.7	0.6	
% transfer tub/shower missing	0.7	0.6	
% locomotion—walk/wheelchair missing	0.9	1.3	
% stairs missing	0.9	0.5	
% comprehension missing	1.3	1.0	
% expression missing	1.4	1.0	
% social interaction missing	1.2	1.6	
% problem solving missing	0.9	0.7	
% memory missing	0.9	0.8	
Matched RICs with motor and cognitive scores			2,959

Rescoring Reliabilities Between Institutional and Calibration Teams on the FIM and the MDS-PAC

An important measure of the performance of an instrument is its reliability. When two individuals or teams use the same assessment tool on a particular case, a reliable tool should yield similar measurements. Measures of rescoring reliability compare how similar the resultant scores are when two individualscomplete the same instrument independently. Here we provide several different statistics for assessing rescoring reliabilities. We compare the institutional team assessments on an instrument with the calibration team assessment using the same instrument on the same patient.

The difference in assessment periods between the FIM and the MDS-PAC were of general concern as a potential source of significant differences. To help understand the role of the assessment period differences, the calibration teams were asked to record two sets of admission FIM data. The first set was to use the same time frame that the institutional team had used. The second set was to be scored as of the end of the reference period. For the inter-rater reliability data reported here, we compare the calibration teams' first set of scores (those scored, albeit sometimes retrospectively, for the same timeframe as that of the institutional teams) with the institutional team scores.

First, we consider Pearson correlation coefficients, which are actually measures of association rather than true measures of agreement. Although there are no absolute standards for what constitutes adequate measures of reliability, McDowell and Newell (1996) suggest that for scales, such as the motor scale and the cognitive scale, a Pearson correlation of .85 and above should be considered acceptable. Both the FIM motor and cognitive scores meet this standard of inter-rater reliability (see Table 4.5).

We know of no agreed-upon standards for item-level correlations. Here, we use a regression framework to help interpret and make judgments about the adequacy of the correlations we observe. In a regression of one variable upon another, the square of the correlation or explained variance equals 1 minus the ratio of the error variance to the total variance. Thus, the proportion of the observed variance of one variable that is explained by another is the square of the correlation. A correlation of [r] for [A] and [B] implies that [r*r] percent of the observed variance in one variable is explained by the other. Unexplained variance accounts for less than half the observed variance in only five of the 13 items. Items from the cognitive scale appear somewhat better, with unexplained variance accounting for 23 to 48 percent of the observed variance.

In the MDS-PAC, several items are combined to create each of the 18 FIM items. We refer to these combined and rescored MDS-PAC items as *pseudo-FIM* items. Table 4.5 indicates that both the pseudo-FIM motor scale and cognitive scales meet the acceptable standard and, in fact, have somewhat higher correlations than the FIM scales. At the item level, eight of the 13 pseudo-FIM motor and four out of five cognitive items have as high or higher correlations than their FIM counterparts. Because the pseudo-FIM items usually combine multiple items from the MDS-PAC, we also looked at correlations for the most important key component within the set of items. These are shown in the rightmost column of

Table 4.5

Rescoring Reliabilities Between Institutional and Calibration Teams: Pearson Correlation Coefficients

	FIM: Institutional Team vs. Calibration Team		PAC: Pseudo FIM Institutional Team vs. Calibration Team		PAC:
FIM Scale or Item	Number	Corre-lation	Number	Corre-lation	Base Item
Motor scale	194	.86	223	.90	
Eating	203	.71	231	.70	.79
Grooming	203	.68	229	.63	.59
Bathing	203	.63	230	.63	.59
Dressing upper body	203	.56	231	.61	.58
Dressing lower body	203	.74	231	.63	.57
Toileting	200	.60	231	.71	.66
Bladder management	199	.57	231	.72	
Bowel management	198	.45	232	.59	
Transfer bed/chair	202	.81	229	.71	.72
Transfer toilet	200	.71	229	.74	.75
Transfer tub/shower	201	.48	228	.58	.57
Locomotion—walk/ wheelchair	202	.68	230	.72	.76
Stairs	201	.71	232	.64	
Cognitive scale	194	.86	224	.87	
Comprehension	203	.80	232	.83	
Expression	203	.88	232	.83	
Social interaction	201	.72	226	.80	
Problem solving	203	.76	211	.85	
Memory	202	.78	231	.84	

Table 4.5. These did not show distinctly stronger or different correlations than the pseudo-FIM items.

Simple and weighted kappa statistics are measures of agreement that correct for chance agreement by calculating the extent of agreement expected by chance alone and removing this from the estimation. In Table 4.6, we present simple and weighted kappas for inter-team scoring reliability for both the FIM and the MDS-PAC's pseudo-FIM items. Simple kappas look at absolute agreement and treat all disagreements the same. Simple kappas are what have been reported most frequently in the literature, so we include them here. Weighted kappas recognize that some disagreements may be more serious than others and weight them accordingly. For example, if team A rated a patient as independent (FIM score of 7) and team B rated the same patient as totally dependent (FIM score of 1), this would be considered a more serious disagreement and thus more heavily weighted than if team B had rated the same patient as modified independent

Table 4.6

Rescoring Reliabilities Between Institutional and Calibration Teams: Kappa Statistics

	FIM		PAC: Pseudo-FIM Items	
FIM Scale or Item	Simple Kappa	Weighted Kappa	Simple Kappa	Weighted Kappa
Motor scale				
Eating	.43	.58	.39	.54
Grooming	.39	.53	.35	.49
Bathing	.45	.55	.33	.46
Dressing upper body	.43	.49	.36	.49
Dressing lower body	.43	.59	.31	.46
Toileting	.40	.51	.27	.51
Bladder management	.25	.43	.40	.60
Bowel management	.18	.31	.28	.47
Transfer bed/chair	.51	.68	.29	.51
Transfer toilet	.48	.61	.36	.57
Transfer tub/shower	.43	.47	.39	.52
Locomotion—walk/ wheelchair	.53	.62	.47	.62
Stairs	.56	.66	.48	.57
Cognitive scale				
Comprehension	.33	.62	.43	.66
Expression	.49	.74	.45	.66
Social interaction	.33	.55	.30	.61
Problem solving	.30	.58	.39	.68
Memory	.33	.61	.38	.67

(FIM score of 6). The weights are proportional to the numeric distance between the response and sum to 1. Because the size of the disagreement will matter in our classification work, the weighted kappa is perhaps the more relevant statistic. Weighted kappas have the same interpretation as simple kappas, so a weighted kappa ≤ .4 represents poor agreement, one between .4 and .6 represents moderate agreement, one between .6 and .75–.8 represents good agreement, and above that agreement is considered excellent (Fleiss, 1981).

For 10 (11) of the 13 motor items, the weighted (simple) kappas on the FIM were greater than or equal to those of the MDS-PAC, suggesting at least somewhat better rating agreement with the FIM motor items than with those on the MDS-PAC. On the cognitive items, the MDS-PAC outperforms the FIM. However, the level of agreement on both instruments is generally only moderate. For the FIM, 8/13 (9/13) motor items have weighted (simple) kappas that demonstrate moderate agreement, whereas 12/13 items (1/13 on the simple kappa) on the PAC are in this range. Four (1) FIM motor items and one (0) PAC pseudo-FIM item have weighted (simple) kappas in the good range. Weighted kappas indicate good agreement for 3/5 FIM cognitive items and 5/5 pseudo-FIM items. Simple kappas indicate much poorer performance.

Although kappa statistics provide information on the quality of agreement, they do not tell us the absolute levels of agreement, nor do they quantify the extent of disagreement. These are shown in Table 4.7 for both the FIM and the MDS-PAC. Absolute agreement at the motor scale level is quite poor—6.7 percent for the MDS-PAC and 7.7 percent for the FIM. It is much better at the item level, ranging from 34.9 percent for bowel management to 88.1 percent for stairs for FIM items and from 39.8 percent for the MDS-PAC's pseudo-FIM toileting item to 80.2 percent for the MDS-PAC stairs item. Absolute inter-team agreement is higher on the FIM than on the pseudo-FIM for 10 of the 13 motor items and two of the five cognitive items. It exceeds 50 percent for 11 of the 13 FIM motor items and for seven of the 13 MDS-PAC pseudo-FIM motor items. However, disagreement (off by more than two points) is also higher for nine of the 13 FIM motor items. Because a substantial proportion of all rehabilitation patients have no measured cognitive impairment, they all score at the maximum level or ceiling. This strong ceiling effect contributes to much higher levels of overall agreement on the cognitive scale—26 percent for the FIM and 31 percent for the MDS-PAC's *pseudo-FIM* cognitive items. The MDS-PAC has higher levels of agreement for three of the five cognitive items and less disagreement for all five items.

Looking across these measures of agreement, we find that despite the lower correlation, the rescoring reliability (as measured by kappas and absolute levels of agreement) of the FIM motor items is modestly better than the MDS-PAC's *pseudo-FIM* items. However, disagreement is also higher, which may explain the lower correlations. On the cognitive scale, the MDS-PAC outperforms the FIM. The FIM's bowel and bladder items have consistently worse rescoring reliabilities. For these items, FIM scoring instructions direct scorers to first evaluate level of assistance with bladder/bowel management on a 1–7 scale and then to evaluate the frequency of accidents on a 1–7 scale to record the lower score, but there is nothing on the scoring to facilitate or remind the scorer of this additional step and more complex scoring. Segal et al. (1993) also found that the bowel and bladder items were the most poorly performing, although Hamilton et al. (1994) and Sharrack et al. (1999) did not. In a meta analysis of FIM reliability studies, Ottenbacher et al. (1996) report that the lowest mean reliability scores among all motor items were for stairs, bladder management, and bowel management. The four motor items with the best scoring reliability in our dataset were stairs, locomotion—walk/wheelchair, transfer bed/chair, and transfer toilet. Although the two transfer items were among the top tertile of performers in the meta analysis, the other two items were in the lowest tertile. The transfer bed/chair item was the most consistent, appearing in the top performance tertile in the three other studies as well as the meta analysis.

Table 4.7

Comparisons of Absolute Agreement Between Institutional and Calibration Teams on the FIM and the MDS-PAC

	FIM: Institutional vs. Calibration Team				MDS-PAC: Pseudo-FIM Institutional vs. Calibration Team			
	Agreement	Within 1	Within 2	Off > 2	Agreement	Within 1	Within 2	Off > 2
Motor scale	7.7	22.7	33.5	66.5	6.7	20.2	29.2	70.9
Eating	60.6	80.8	95.1	4.9	61.0	71.0	97.8	2.2
Grooming	63.6	84.7	96.1	3.9	63.3	78.6	95.6	4.4
Bathing	58.6	88.2	95.6	4.4	47.8	83.9	97.8	2.2
Dressing upper body	60.6	79.8	89.2	10.8	55.0	75.8	90.5	9.5
Dressing lower body	54.7	86.7	97.0	3.0	43.7	83.6	93.1	6.9
Toileting	51.5	76.5	93.0	7.0	39.8	81.4	93.5	6.5
Bladder management	37.7	57.8	73.4	26.6	51.5	72.3	85.3	14.7
Bowel management	34.9	61.6	82.3	17.7	43.5	80.6	85.8	14.2
Transfer bed/chair	64.4	92.6	98.5	1.5	44.5	85.2	96.9	3.1
Transfer toilet	61.5	88.0	94.0	6.0	49.3	87.3	96.9	3.1
Transfer tub/shower	72.3	80.7	84.2	15.8	64.9	78.5	88.2	11.8
Locomotion—walk/wheelchair	66.3	83.7	91.1	8.9	63.0	87.4	91.3	8.7
Stairs	88.1	95.5	97.0	3.0	80.2	88.8	91.4	8.6
Cognitive scale	26.0	44.5	65.5	34.5	31.3	49.1	603	39.7
Comprehension	56.2	88.2	96.1	3.9	60.3	91.4	98.7	1.3
Expression	69.5	91.1	98.5	1.5	61.6	91.8	98.7	1.3
Social interaction	57.2	85.1	94.5	5.5	48.2	85.8	96.5	3.5
Problem solving	46.3	75.5	93.1	6.9	53.5	84.4	97.4	2.6
Memory	49.5	79.7	94.6	5.5	52.8	86.6	96.5	3.5

Are the FIM and the MDS-PAC Measuring the Same Concepts?

Factor analysis is a set of techniques that help analysts synthesize complex datasets into constructs or factors that enable us to better observe relationships among variables. It is particularly useful for comparing variables and scales to determine the extent to which they are measuring the same concepts. In this project, we compare two distinctly different instruments that seek to measure many common elements. We have developed translations that combine variables from the MDS-PAC to make them more like those in the FIM. Factor analysis is a tool that allows us to measure whether we were successful in doing so. We use factor analysis to help us understand (1) how items within the MDS-PAC cluster to form common factors, (2) whether the combined FIM and the MDS-PAC raw items load onto common or distinct factors (this tells us the extent to which the basic instruments are measuring the same or different constructs), (3) how well the translated or pseudo-FIM items constructed from raw MDS-PAC items load together with the actual FIM items (which shows us if the translations do better than the raw items at measuring similar constructs), and (4) how different pseudo-FIM translations compare (did the work presented here on the new motor translation actually improve the comparability of the underlying constructs?). We also used cluster analysis to see how the FIM and pseudo-FIM items clustered.

Using PROC FACTOR in SAS software, we applied maximum likelihood (ML) factor analysis to the MDS-PAC data to examine how various concepts are explained by each section of the MDS-PAC instrument. Varimax factor rotation was applied to the factors we obtained from the ML factor analysis. Varimax rotation separates loadings within a factor—that is, loadings on a rotated factor take on more extreme (higher and lower) values than those of an unrotated factor; this aids with factor interpretation, since it clarifies which items load highly on which factors. We set the prior communality estimates equal to their squared multiple correlations with all other variables, which is the default method in SAS. Most often, the number of factors that were retained in a factor analysis explained 100 percent of the common variance (i.e., the variance explained by the common factors) using the prior communality estimates. When there was a small number of items relative to the possible number of factors, selection of the number of factors in this way was often not possible, in which case we retained the maximum number of factors allowed. In our analysis of the crosswalk pseudo-FIM versus the FIM scores, we also conducted a variable cluster analysis to see whether there was agreement between those results and the factor analysis. The variable clusters are disjoint, which differs from the factor analysis because the factors “overlap”—all variables load somehow onto all factors. We used PROC VARCLUS in SAS software.

How Items Within the MDS-PAC Cluster to Form Common Factors

First, we analyzed sections B through F of the MDS-PAC. We performed separate unrotated factor analyses on the items in MDS-PAC sections B through F to determine how many factors each section contains. For each analysis, we used the proportion criterion in SAS PROC FACTOR to decide how many factors to retain. In other words, we retained the factors that had the largest eigenvalues and accounted for 100 percent of the common variance. Examination of the eigenvalues (Table 4.8) reveals that one dominant factor emerges in each of these MDS-PAC sections; the proportion of the common variance that is explained by the dominant factor in each section is given in the last row of Table 4.8. With the exception of section MDS-PAC-E, these results suggest that the items in these

Table 4.8

Eigenvalues of Factors for the Unrotated Factor Analyses of Each MDS-PAC Section

Factor	PAC-B Cognitive Patterns	PAC-C Communication/ Vision	PAC-D Mood and Behavior	PAC-E Functional Status	PAC-F Bladder/ Bowel Management
1	28.83	158.54	40.02	242.19	34.76
2	2.59	7.89	3.84	45.51	3.84
3	0.52	0.77	2.53	17.73	2.43
4	0.25	0.69	0.60	10.70	0.65
5	0.09	0.18	0.31	7.93	0.39
6	0.03	0.09	0.25	6.58	0.28
7	0.02	–0.01	0.16	4.37	0.15
8	–0.03	–0.08	0.03	2.51	0.09
9	–0.06	–0.17	0.00	2.03	0.04
10	–0.13	–0.22	–0.05	1.12	–0.01
11	–0.14	–0.33	–0.05	0.99	–0.07
12	–0.24	–0.36	–0.06	0.97	–0.09
13	–0.31	–0.57	–0.09	0.84	–0.23
14	N/A	N/A	–0.20	0.67	–0.31
15	N/A	N/A	–0.23	0.60	–0.43
16	N/A	N/A	–0.30	0.52	–0.46
17	N/A	N/A	–0.35	0.44	N/A
Number of factors retained	2	2	3	9	3
Percentage of the common variance explained by factor 1	91.8	95.3	86.3	71.3	84.7

NOTE: MDS-PAC-E had 59 variables, so eigenvalues corresponding to just the top 17 factors are presented. N/A: does not apply because the maximum number of factors have been found.

MDS-PAC sections highly load onto one factor, representing one construct per section. The first factor of MDS-PAC-E explains 71.3 percent of the variation; although this is considerable, the second factor explains 13.4 percent of the variation, with the remaining factors representing the rest. Thus, it is likely that more than one construct is distinctly represented by MDS-PAC-E; device-related items load on the first unrotated factor and gross motor items (transfers, locomotion, dressing, hygiene) load on the second factor. The number of factors retained for further analysis in each MDS-PAC section is also given in Table 4.8.

Next, we rotated the factors that were retained in the unrotated factor analysis to further explore whether various constructs are represented by each MDS-PAC section. Rotation clarifies the interpretation of factors by separating the loadings within a factor, so that loadings have more extreme values relative to unrotated factors. The resulting rotated factors under varimax rotation are uncorrelated, which makes it likely that a given variable will load heavily on just one or a small number of factors rather than on a greater number of factors, which facilitates interpretation of the factors. An oblique (promax) rotation, which allows for correlated factors, was also considered. The factors obtained using both promax and varimax rotations agreed for these sections, so just the rotated factors that were obtained by using a varimax rotation are presented here.

Table 4.9 reports the number of factors found for each section, a description of each factor, and the proportion of the common variance explained by each factor. The common variance is the variation in the data that is explained by the factors; it does not include variation resulting from factors or errors that are specific to a variable. The proportion of the variance explained by the common factors is estimated in two ways. The unweighted estimates treat all the variables equally in the analysis. The second method weights each variable by the reciprocal of its variance that is unexplained by the common factors. Thus, variables that have a greater explanatory power are more heavily weighted in this calculation than other variables. The weighted results can be regarded as approximating a factor analysis that retains just those variables found to have greater explanatory power than other variables. We will focus on the unweighted results in the discussion below, but these can be compared to the weighted estimates to shed light on the explanatory power of certain subsets of variables in each analysis. Rotated factor loadings and communalities for items in MDS-PAC sections B through F appear in Tables 1A–4J of Appendix J. Because of the large number of items in MDS-PAC-E, only those variables with primary factor loadings (i.e., those with absolute value greater than 0.3) are presented for each factor in Tables 4A–4I of Appendix J.

Table 4.9

Factors in Sections B–F of the MDS-PAC

Section or Factor	Number of Factors	Unweighted Proportion of Variance Explained	Weighted Proportion of Variance Explained
Section B. cognitive patterns	2		
Awareness/orientation		52.6	50.1
Memory		47.4	49.9
Section C. communication patterns	2		
Modes of communication		56.8	87.4
Clarity and understanding		43.2	12.6
Section D. mood and behavior	3		
Abusive behavior		44.9	57.1
Fear and negativity		36.2	28.6
Sadness and withdrawal		18.9	14.4
Section E. functional status	9		
Devices		23.9	63.4
Gross motor: transfers, locomotion, self-care (toileting, dressing lower body, bathing)		23.3	9.0
ADL assistance		18.7	9.5
Impairment mobility and motor control		13.6	5.8
Instrumental—managing finances, phone use, medication management		8.2	3.3
Number of of ADLs previously independent and number more limited now		4.7	4.2
Distance and walking support		3.9	3.3
Stairs and car transfers		1.9	0.7
ADL assistance (grooming, eating, dressing upper body)		1.9	0.9
Section F. bowel and bladder management	3		
Bladder appliance		49.6	61.7
Bowel appliance		29.7	27.5
Bowel and bladder continence and appliance support		20.7	10.7

Nine factors were identified in the functional status section using the proportion criterion described above. However, examination of factors 8 and 9 (Tables 4H and 4I of Appendix J) shows only a few items each loading for each factor, and the loadings are fairly low, all being less than 0.45. Further, the common variance explained by either of these two factors is less than 2 percent in both the weighted and unweighted estimates, so they could be removed from further consideration. As is clear from Table 4.9, the most important factor to emerge is that for devices (Table 4A, Appendix J), which explains 23.9 percent of the common variance. A gross motor (transfers, locomotion, self-care (toileting, dressing lower body, bathing)) factor that explains 23.3 percent of the common variance was identified (Table 4B of Appendix J). ADL Assistance Codes (one limb or two or more person assistance) explained another 18.7 percent of the common variance. Similarly, in the bowel and bladder management section, three factors were identified, (1) bladder appliances, (2) bowel appliances, and (3) bowel and bladder continence and appliance support. Within this section, bladder appliance explained 49.6 percent of the variance, bowel appliance explained another 29.7 percent, and the third factor explained the remaining 20.7 percent.

Our literature review did not reveal any factor analytic work on the MDS-PAC. However, Casten et al. (1998) used confirmatory factor analysis on the nursing home MDS Activities of Daily Living (bed mobility, transfer, locomotion, dressing, eating, toilet use, personal hygiene, bathing, bowel and bladder continence). They found that in their exploratory and confirmatory samples, the 10 items appear to form a compact factor with high internal consistency but poor fit statistics. When they compared a cognitively intact subsample with a cognitively impaired group, they found that the two groups had markedly different structures. In particular, in the impaired group, the two types of incontinence had highly correlated residuals and contributed substantially to the lack of fit. They also found that the impaired group had a high unique shared variance between gross motor items (bed mobility, transfer, and general locomotion) that was not found in the intact group. These gross motor items are similar to those in our second PAC functional status factor but the former do not include toileting, bathing, and dressing.

How the Combined FIM and the MDS-PAC Raw Items Load onto Common or Distinct Factors

To see how the MDS-PAC items worked with the FIM items, we combined all items from the relevant sections of the MDS-PAC and the FIM and then separated the cognitive and the motor items. In the analysis of cognitive items, we included the five FIM items (comprehension, problem solving, expression, memory, and social interaction) with all MDS-PAC items from Section B (cognitive patterns), Section C (communication patterns) and Section D (mood and behavior), along with one item from Section K (K5C: trunk restraint) and Section L (L2C: patient fails to initiate or continue doing ADLs (once initiated) for which he/she has some demonstrated capability); these two latter items are part of the Morris translation. The five FIM items have good communalities with a low of .7 for social interaction up to .92 for problem solving. Each communality is the portion of the variance of each variable that is explained by the set of common factors identified by the factor analysis. The top 20 (of 50) eigenvalues obtained from an unrotated factor analysis are listed in Table 4.10; four of the factors obtained have notably larger eigenvalues than the rest. The proportion criterion in SAS identified eight factors. Eight factors—essentially the seven factors associated with MDS-PAC Sections B, C, and D along with K5C and L2C above and an eighth factor for the FIM cognitive scale—were identified in this analysis. The FIM cognitive items remain together (factor 2, Table 6B, Appendix J) contributing almost nothing to the other seven factors. The FIM memory item does not load meaningfully on the MDS-PAC memory factor nor does the FIM social interaction load on the MDS-PAC reduced social interaction factor. Similarly, the FIM expression item does not load on the MDS-PAC expression factor. The second factor, the FIM cognitive scale (the five FIM items), accounts for only 19.4 percent of the explained variance (see Table 4.11). The eighth factor identified could be dropped, since only two items loaded at .30 (Table 6H of Appendix J) and because those two items had higher loadings on another factor (factor 2). The MDS-PAC communication mode items have much greater portion of common variance than the other items, as reflected by the high weighted percentage of common variance explained of 53.9 percent.

Table 4.10

Top 20 Eigenvalues for Unrotated Factor Analysis of Combined FIM and Raw MDS-PAC Cognitive Items

Factor	Eigenvalue
1	153.86
2	51.25
3	28.25
4	7.35
5	3.41
6	3.21
7	2.46
8	2.32
9	1.17
10	0.84
11	0.61
12	0.45
13	0.42
14	0.38
15	0.28
16	0.21
17	0.21
18	0.17
19	0.14
20	0.12

Table 4.11

Factors Resulting When MDS-PAC and FIM Cognitive Items Are Combined

Factor	Unweighted Percentage of Variance Explained	Weighted Percentage of Variance Explained
Anger, fear, and anxiety	22.9	10.2
FIM cognitive scale	19.4	16.1
Communication modes	17.0	53.9
Awareness/disordered thinking	16.1	6.8
Memory	12.5	5.7
Abusive behavior	6.1	4.2
Withdrawal/reduced social interaction	4.5	1.7
FIM memory and FIM problem solving	1.7	1.4

There is a much closer item-by-item correspondence (albeit with reversed scaling) between the FIM motor items and a subset of the MDS-PAC functional status and bowel and bladder management items (Sections E and F of MDS-PAC and items A–M of FIM). In this factor analysis of motor items, we found that these behave differently than the cognitive items. The communalities for transfer tub/shower and stairs are below .5 on both instruments, whereas bladder management and bathing are below this level on the MDS-PAC but not on the corresponding FIM items (see Table 4.12). The bowel and bladder management items on both instruments combine two concepts, continence and management of the function. On the FIM, raters are instructed to score each but record only the lower score. On the MDS-PAC, these appear as two separate items with somewhat altered definitions, so we report two communalities.

Table 4.12

Communalities on Raw MDS-PAC and FIM Functional Status Items

Item	FIM	PAC
Dressing upper body	0.73	0.70
Bladder management	0.75	0.27/0.44
Dressing lower body	0.72	0.60
Transfer bed/chair	0.64	0.78
Grooming	0.60	0.63
Toileting	0.60	0.66
Bowel management	0.56	0.42/0.23
Transfer toilet	0.56	0.73
Eating	0.57	0.57
Bathing	0.56	0.41
Transfer tub/shower	0.39	0.17
Locomotion—walk/wheelchair	0.39	0.46
Stairs	0.35	0.33

Using the proportion criterion in SAS, 13 unrotated factors were identified across the combined functional status and bowel and bladder management items; the top 20 (of 88) eigenvalues for the factors are presented in Table 4.13. The primary and secondary loadings for 13 rotated factors are presented in Tables 7A–7M in Appendix J, and the factors are summarized in Table 4.14. The first factor comprises MDS-PAC gross motor items such as transfer bed/chair, transfer toilet, instrumental ADL stairs, bed mobility, instrumental car transfer, and toilet use. The similar FIM items load negatively (because of the reverse scoring) as one would expect. Factors 2, 3, 4, and 6 (assistive devices, bladder appliances, levels of physical assistance on ADLs, and impairments) are not specifically articulated in the FIM and the FIM items do not load on these factors in substantive ways. Factor 5 is the instrumental ADLs that require cognitive competence (managing money, phone use, medication management) and the fine motor ADLs (eating, grooming, dressing upper body). The fine motor ADLs in the MDS-PAC load on this factor (the FIM does not include the instrumental ADLs). Factors 7 and 8 are identified by the FIM motor items. The 11 self-care and mobility/locomotion items form one factor and the bowel and bladder management items the second factor. The corresponding MDS-PAC items load with small negative correlations. Ravaud et al. (1999) also used factor analysis to

Table 4.13

Top 20 Eigenvalues for Unrotated Factor Analysis of Combined FIM and Raw MDS-PAC Motor Scale Items

Factor	Eigenvalues
1	206.08
2	55.57
3	26.72
4	14.30
5	12.56
6	9.85
7	7.04
8	5.51
9	3.63
10	3.24
11	3.01
12	2.49
13	1.85
14	1.24
15	1.21
16	1.11
17	.97
18	.89
19	.86
20	.78

Table 4.14

Thirteen Factors Identified in the Analysis of Combined FIM and MDS-PAC Motor Scale Items

Factor	Unweighted Percentage of Variance Explained	Weighted Percentage of Variance Explained
1. MDS-PAC gross motor—transfers, instrumental stairs and car transfers, bed mobility	20.1	9.7
2. MDS-PAC assistive devices	16.7	52.5
3. MDS-PAC bladder appliances	13.5	9.5
4. MDS-PAC ADL assistance levels—self-care items	12.2	6.5
5. MDS-PAC instrumental ADLs, plus grooming, eating, and dressing upper body	9.2	4.3
6. MDS-PAC impairments	9.1	4.2
7. FIM mobility and self-care items	5.1	2.2
8. FIM bladder and bowel management	3.4	1.4
9. MDS-PAC number of ADLs independent previously and number now limited in	3.3	4.5
10. MDS-PAC walking support and farthest distance walked	2.5	2.8
11. MDS-PAC bowel appliance	1.8	1.0
12. MDS-PAC functional status (dressing, grooming, eating)	1.6	0.7
13. MDS-PAC ADL assistance—grooming, dressing, eating	1.5	0.8

study the FIM and found that an orthogonal transformation supports the use of the 13-item motor scale and the five-item cognitive scale but with a Varimax rotation the motor scale decomposes into three factors—one for gross motor items, another for self-care, and a third for bowel and bladder management. Interestingly, toilet use loaded more with the motor items as we observed in the PAC. Stineman et al. (1997a) also looked at the FIM using factor analysis, finding that motor scale formed one, two (mobility and self-care), or three factors (mobility, self-care, and bowel/bladder) depending upon the RIC. Depending upon the RIC, the toilet use item grouped into each of the three factors. With respect to the proportion of the variance explained, the most important factors are listed in order in Table 4.14; namely, the MDS-PAC gross motor items explain 20.1 percent of the common variance, whereas MDS-PAC assistive devices, MDS-PAC bladder appliances, and MDS-PAC ADL assistive devices explain 16.7 percent, 13.5 percent, and 12.2 percent of the common variance, respectively, with the proportion of common variance explained for the remaining factors given in the table. Using weighted estimates, the MDS-PAC assistive devices item explains far more of the common variance than do the other items (52.5 percent), with gross motor (9.7 percent), and bladder appliances (9.5 percent) following; thus, the MDS-PAC assistive items have more explanatory power than the other items.

Comparing Translations of MDS-PAC-Based Items with FIM

Next, we combined information across items in the MDS-PAC to create new items that more closely approximated the 18 FIM motor and cognitive scale items. We benefited from earlier research conducted by Dr. John Morris who provided an initial set of item "translations" from MDS-PAC items to the FIM. (The Morris translations are described in Appendix K.) The study team took these as a starting point and further refined the motor scale item translations (see Section 3). We continue to use the Morris translations of cognitive items and refer to our translations as pseudo-FIM items. In addition to using factor analysis in our third set of analyses, we also used cluster analysis as a check to compare the 18 motor and cognitive scale items from the FIM with the 18 pseudo-FIM items from the MDS-PAC. We also compare the Morris and study translations of motor items.

Eigenvalues for the top 10 of 36 unrotated common factors identified in the analysis of 18 pseudo-FIM and FIM items are presented in Table 4.15. Eight common factors were identified by the proportion criterion in SAS. The communalities were generally high with the exception of transfer tub/shower, stairs, and sphincter control items, which have communalities of .5 or less (Table 8I of Appendix J). The corresponding FIM and pseudo-FIM items have similar communalities. The eight factors along with the percentage of common variance each explains are in Table 4.16, and the primary and secondary factor loadings are

Table 4.15

Top 10 Eigenvalues for Unrotated Factors in the Analysis of Pseudo-FIM and FIM Item Scores

Factor	Eigenvalue
1	147.63
2	24.55
3	12.56
4	5.46
5	3.91
6	3.14
7	2.23
8	1.51
9	0.89
10	0.73

Table 4.16

Eight Factors Identified in the Analysis of Pseudo-FIM and FIM Scores

Factor	Unweighted Percentage of Variance Explained	Weighted Percentage of Variance Explained
1. Cognitive scale items (comprehension, expression, problem solving, memory, and social interaction	35.2	62.0
2. Dressing, bathing, toileting, and transfers	21.2	9.1
3. Locomotion and transfers	13.7	4.5
4. Eating, grooming, dressing upper body	10.4	5.7
5. Sphincter control (bladder, bowel management)	9.7	5.2
6. Transfers (toilet and bed) and toileting	5.7	3.8
7. FIM expression and comprehension	2.3	5.9
8. FIM problem solving and memory	1.9	3.8

in Tables 8A-8H in Appendix J. With the exception of factors 7 and 8, the other factors contain both the MDS-PAC and FIM items on corresponding measures.

In addition to the factor analysis, we also looked at these 36 items using cluster analysis. Unlike factor analysis, cluster analysis restricts each variable to belong to only one cluster, whereas each variable contributes to each factor in the factor analysis. Clusters are selected as follows: cluster components are computed as averages of standardized variables, and then the variance of the original variables that is explained by the cluster components is maximized. Six clusters were identified. Table 4.17 shows how many variables belong to each cluster, the variation explained, and the proportion of variance attributable to the variables of a cluster that is explained by the cluster designation. The items in each cluster, the correlation of that variable with the cluster component, the next highest squared correlation of that variable with another cluster component, and the ratio of (1 – own cluster R**2) to (1 – next cluster R**2) are found in Table 4.18. This last column measures the degree of separation of the clusters with lower values indicating more separation.

Table 4.17

Cluster Analysis Output for Combined Motor and Cognitive Items

Cluster	Members	Variation Explained	Proportion Explained
1	10	8.07	0.81
2	10	6.37	0.64
3	6	4.02	0.67
4	4	2.49	0.62
5	4	2.29	0.57
6	2	1.54	0.77

Table 4.18

Cluster Definitions

Variable	Own Cluster Correlate	Next Closest Correlate	1 – R**2 Ratio
Cluster 1			
COMREHD	0.87	0.35	0.20
EXPRESS	0.85	0.35	0.23
SOCIAL	0.83	0.32	0.25
PROBSOLV	0.83	0.39	0.27
MEMORY	0.87	0.32	0.19
FIM COMPREHD	0.76	0.27	0.33
FIM EXPRESS	0.75	0.31	0.36
FIM SOCIAL	0.69	0.30	0.44
FIM PROBSOLV	0.80	0.30	0.29
FIM MEMORY	0.81	0.30	0.28
Cluster 2			
BATHING	0.56	0.32	0.64
DR_LOWER	0.65	0.31	0.51
TOILET	0.70	0.41	0.50
TR_BED	0.73	0.34	0.41
TR_TOILT	0.75	0.38	0.41
FIM BATHING	0.51	0.29	0.70
FIM DR_LOWER	0.62	0.31	0.54
FIM TOILET	0.60	0.36	0.62
FIM TR_BED	0.67	0.30	0.46
FIM TR_TOILT	0.57	0.24	0.56
Cluster 3			
EATING	0.64	0.29	0.50
GROOMING	0.67	0.36	0.52
DR_UPPER	0.69	0.48	0.59
FIM EATING	0.67	0.31	0.48
FIM GROOMING	0.68	0.30	0.46
FIM DR_UPPER	0.67	0.42	0.57
Cluster 4			
BLADDER	0.63	0.25	0.50
BOWEL	0.63	0.25	0.63
FIM BLADDER	0.72	0.24	0.37
FIM BOWEL	0.61	0.20	0.48
Cluster 5			
WALK_WC	0.66	0.33	0.51
STAIRS	0.64	0.15	0.54
FIM WALK_WC	0.60	0.25	0.54
FIM STAIRS	0.50	0.10	0.56
Cluster 6			
TR_TUB	0.77	0.16	0.27
FIM TR_TUB	0.77	0.18	0.28

Overall, the six clusters explain almost 69 percent of the total variation in the data. All the clusters contain both the FIM and the corresponding pseudo-FIM item for the variables within them. Cluster 1 contains the five cognitive scale items from the FIM and the five corresponding pseudo-FIM items from the MDS-PAC. Surprisingly, this cluster has the highest level of total variation explained (8.07/36.00 = 0.22). The second cluster contains two transfer items (toilet and bed/chair), along with toileting, dressing lower body, and bathing from both the FIM and the pseudo-FIM MDS-PAC items. Eating, grooming, and dressing upper body form the third cluster. The sphincter control items (bladder and bowel management) form cluster four. The two locomotion items, walk-wheelchair and stairs, make up cluster five. Interestingly, transfer tub/shower is its own cluster, which suggests that it operates differently from the others. Transfer tub/shower also had the lowest communality in the factor analysis, further supporting this notion. This is also consistent with the RAND team findings, which noted that transfer tub/shower did not have the expected relationship to costs. The RAND team has now recommended dropping this item from the motor scale when the new CMGs are created. The clusters are well-separated; the first cluster (FIM cognitive items and corresponding MDS-PAC items) has low values in the 1 – R**2 ratio column, and the transfer tub/shower cluster has low ratios, too, reinforcing its separation from the other items.

Last, we compare the performance of our pseudo-FIM translations for the motor scale items with that of those recommended by Dr. John Morris. Both the factor analysis and the cluster analysis support our translation over the Morris version. Six common factors were found for the pseudo-FIM and FIM items, whose largest eigenvalues are presented in Table 4.19. Five common factors were found for the Morris translation and the FIM items, based on the unrotated factors; the eigenvalues for all of the unrotated factors are in Table 4.20. The rotated factors for both analyses are summarized in Tables 4.21 and 4.22; the primary and secondary factor loadings for these factors are given in Appendix J for the pseudo-FIM and FIM items (Tables 9A–9F) and for the Morris translation and FIM analysis (Tables 10A–10E). Our translations produce six factors, five of which contain both the corresponding FIM and pseudo-FIM items and identify several important concepts. The factor analysis of the Morris translation and the FIM scores produced five factors, only two of which corresponded to FIM and MDS-PAC items simultaneously.

We also found that the factor loadings of the corresponding pseudo-FIM and FIM items were more similar to our translations than to those of the Morris versions. The proportion of the total variation explained in the variable cluster analysis of motor items was quite similar, .64 for our translation and .63 for the Morris translation. The clusters were very similar as well with one exception. In the Morris translation, the locomotion walk/wheelchair item from the MDS-PAC falls into the transfer, toileting, dressing lower body, and bathing cluster, leaving the FIM locomotion item alone in a cluster with the FIM stairs item and the pseudo-FIM stairs item.

Table 4.19

Top 10 (of 26) Eigenvalues for Unrotated Factors of Pseudo-FIM and FIM Motor Items

Factor	Eigenvalue	Factor	Eigenvalue
1	37.26	6	1.22
2	5.56	7	0.72
3	3.30	8	0.58
4	2.48	9	0.41
5	1.66	10	0.34

Table 4.20

Top 10 Eigenvalues for Unrotated Factors of the Morris Translation and FIM Motor Items

Factor	Eigenvalue	Factor	Eigenvalue
1	36.02	6	0.81
2	6.35	7	0.62
3	4.61	8	0.54
4	2.33	9	0.37
5	1.31	10	0.33

Table 4.21

Six Factors Derived from the Factor Analysis of Pseudo-FIM and FIM Motor Scores

Factor	Unweighted Percentage of Variance Explained	Weighted Percentage of Variance Explained
1. Eating and grooming	20.2	21.4
2. Locomotion—walk/wheelchair and stairs	17.2	13.2
3. Bladder and bowel management	17.1	17.0
4. Transfers and toileting	17.0	22.9
5. Dressing	16.5	15.2
6. FIM self-care	12.1	10.3

Table 4.22

Five Factors Derived from the Factor Analysis of the Morris Translation and FIM Motor Scores

Factor	Unweighted Percentage of Variance Explained	Weighted Percentage of Variance Explained
1. Eating and grooming	23.6	22.8
2. MDS-PAC transfers	23.2	17.6
3. FIM bladder and bowel management	21.1	22.9
4. MDS-PAC bladder and bowel management	16.1	22.8
5. Dressing	16.0	13.9

Instrument Completion Times

An important part of the evaluation of any instrument is the administrative burden it places on the institutions using it. The widespread use and voluntary acceptance of the FIM comes in part because its designers were seriously concerned with minimizing its administrative burden. The MDS-PAC designers had different goals, wanting an instrument that could be used for a broader set of purposes and could also be used to assess patients possibly at different stages of illness and recovery across a variety of care settings. To assess instrument burden, we asked all data collectors (institution and calibration teams) to record on the face sheet of the instrument the amount of time they spent collecting data. Data collectors were instructed not to include actual clinical assessment time. We did *not* provide stopwatches. We did record how many people contributed to each assessment and the clinical credentials of the assessors. During our training sessions, we focused on how to complete the instrument correctly; we did not discuss how to complete it efficiently. We asked institutions to send and use multidisciplinary data collection teams. We expected each team to include at least a registered nurse, a physical therapist, and an occupational therapist. The participation of speech language pathologists was encouraged but not required. However, institutions did not always use teams to collect data and, in fact, we found that nearly a quarter of our cases had only one data collector.

Our findings on mean overall completion times by instrument are shown in Table 4.23. The average time to complete the admission FIM by all institutional teams was 25 minutes. Because our calibration teams were unfamiliar with each institution's medical records and procedures and had no familiarity with the patients, they took notably longer, 148 minutes per case. The average time for institutional teams to complete the MDS-PAC was 147 minutes, almost 2-1/2 hours. We winsorized the data, deleting the top 5 percent and the bottom 5 percent and found that this reduced the mean time by only nine minutes, indicating that a small percentage of large outliers are not driving this mean. MDS-PAC completion times for the calibration teams were again notably longer, with a mean time of 221 minutes (over 3-1/2 hours). Our data demonstrate a learning curve; average times for institutional teams during the first two weeks were 184 minutes (just over three hours) falling to 120 minutes (two hours) by weeks 7 and 8. We also found that the size of the data collection team had a significant effect on the data collection time; larger teams spent more time. When the MDS-PAC was completed by one person, completion times averaged 113 minutes. Four or more person teams spent an average 175 minutes.

Table 4.23

Average Time Required to Complete Each Instrument (Standard Deviations) (in minutes)

	Institution Teams: 3,409 Cases		Calibration Teams: 241 Cases	
	FIM	PAC	FIM	PAC
Overall	25 (19)	147 (75)	148 (108)	221 (75)
Period				
Weeks 1–2	27 (21)	184 (88)	193 (111)	259 (86)
Weeks 3–4	24 (19)	149 (69)	145 (103)	228 (66)
Weeks 5–6	24 (19)	133 (64)	134 (112)	208 (73)
Weeks 7–8	23 (18)	120 (57)	121 (95)	189 (54)
Team size				
1 person	17 (12)	113 (42)		
2 people	23 (12)	147 (69)		
3 people	22 (18)	166 (85)		
4+ people	35 (23)	175 (85)		

Next we looked for a learning curve effect within team size (see Table 4.24) and we found one. For one-person teams, the mean completion time fell from 140 minutes during the first two weeks to 91 minutes during weeks 7 and 8. Similarly, for each team size, times fell 50 to 74 minutes across the eight weeks. We note that the average completion time that we found for the one-person teams was similar to what Dr. Morris reported (85 minutes) for experienced data collectors as referenced in the Notice of Proposed Rule Making. Presumably, multidisciplinary teams should produce more accurate information, but our data indicate that this accuracy may come with a fairly high cost in terms of the additional administrative burden it represents.

Finally, we wanted to look at the effect of other variables on time to complete the MDS-PAC. We used a regression framework and transformed completion times by taking the natural logarithm to reduce the skewness. Then T_{ij} is the log (time required to complete the MDS-PAC for patient i at hospital j) and our model is specified as:

$$T_{ij} = \alpha + X_{ij}\beta + Y_{ij}\delta + Z_j\gamma + \eta_j + \varepsilon_{ij}$$

The vector X_{ij} represents the characteristics of form administration which include the timing of administration (period 1 is the first one-third of the study, period 2 is the middle one-third, and period 3 is the end of the study timeframe) and the number of data collectors (one person, two people, three people, four people, or five or more persons). The vector Y_{ij} denotes a set of patient characteristics and the vector Z_j a set of hospital characteristics. The set of patient characteristics includes age (in four groups), gender, RIC (stroke, lower extremity fracture,

Table 4.24

MDS-PAC Completion Times by Period and Team Size (Standard Deviations) (in minutes)

Period	1 Person	2 People	3 People	4+ People
Weeks 1–2	140 (46)	180 (76)	204 (90)	217 (106)
Weeks 3–4	120 (36)	149 (68)	172 (91)	167 (67)
Weeks 5–6	107 (39)	139 (67)	139 (68)	161 (75)
Weeks 7–8	91 (31)	127 (55)	142 (72)	143 (57)
Sample size	827	680	703	1,197

lower extremity joint replacement, all other RICs), functional status group (3 = motor score ≤ 41; 2 = motor score ≤ 53; 1 = motor score > 53), poor cognitive status (cognitive score ≤ 30), and poor ability to communicate (impaired hearing, difficulty making oneself understood, or difficulty understanding others). The vector of hospital characteristics includes size (as measured by the average daily census reported for June 2000), urban (versus rural) setting, and freestanding facilities. The η_j represents a random effect for hospital. This is included to test whether completion times cluster within facilities after controlling for other characteristics.

The results are shown in Table 4.25. The coefficients are all relative to an omitted group that includes cases done in period 3, scored by four or more people, on patients with good cognition, high motor function, good ability to communicate, in a small rural hospital. As expected, the administrative variables of timing and team size were highly significant. We found that gender and inability to communicate had no effect. Lower extremity joint replacement cases took less time. Patients with lower functional status and those with poor cognition required significantly more time. None of the hospital characteristics mattered. The random effect for hospitals was significant.

Summary

In this section, we have analyzed several aspects of instrument performance including ease of completion, inter-rater reliabilities, conceptual breadth and cohesion, and administrative burden. Important insights include (1) without menu driven software for impairment code selection, data collectors had difficulty with this item; (2) substantial disagreement in impairment code selection was observed between calibration teams and institutional teams indicating ambiguity and the need for better written guidance on impairment code selection; (3) inter-team scoring reliabilities as measured by kappa statistics and absolute agreement were somewhat higher for the FIM than the MDS-PAC,

Table 4.25

Regression Model Explaining Log (MDS-PAC Completion Time)

Variable	Parameter Estimate	Standard Error	p-Value
Intercept	4.86	0.11	0.0001
Period 1	0.38	0.01	0.0001
Period 2	0.15	0.01	0.0001
1-person scoring team	–0.33	0.03	0.0001
2-person scoring team	–0.26	0.02	0.0001
3-person scoring team	–0.09	0.02	0.0001
Age < 65	0.07	0.02	0.0008
65 < age < 75	0.01	0.01	0.3362
75 < age < 85	0.03	0.01	0.0313
Low motor score	0.10	0.01	0.0001
Medium motor score	0.06	0.01	0.0001
Low cognitive score	0.06	0.01	0.0001
Poor communication	0.03	0.02	0.1556
Male	0.00	0.01	0.8908
Lower extremity joint replacement RIC	–0.06	0.01	0.0001
Small hospital	–0.12	0.14	0.3678
Medium hospital	–0.06	0.11	0.5732
Rural hospital	0.09	0.13	0.4833
Freestanding hospital	–0.15	0.11	0.1746

but were not as high as previously reported in the literature or as desired for use in a payment system; (4) the greater conceptual breadth of the MDS-PAC motor and cognitive items is clearly demonstrated in the factor analysis, which also showed little overlap in raw factors measured by the two instruments; (5) translations that combined several MDS-PAC items into pseudo-FIM items created items that loaded onto factors in a similar fashion and clustered together; (6) factor analysis indicated lower communalities for transfer tub/shower, sphincter control items, and stairs, suggesting that these operate differently than the other items; cluster analysis findings, which put transfer tub/shower in its own single item cluster and the two sphincter control items in another support this; (7) the study MDS-PAC translations into pseudo-FIM items improve upon those previously proposed by Dr. Morris; and (8) the administrative burden of the MDS-PAC overall is nearly six times greater than that of the FIM. There is a clear learning curve effect, which may continue to bring the times down from those reported here. The size of the data collection team also influences data completion times significantly. Poor motor function and poor cognition in patients increase completion times. Hospital characteristics have no effect but hospitals differ significantly from one to the next in their completion times after controlling for patient and administrative characteristics.

5. Accuracy of the MDS-PAC Translation into Pseudo-FIM Items

Introduction

To evaluate the effect of the new motor item translations, we recomputed scale and item-level means and these are shown in Table 5.1. This was done for both the institutional assessments (~3,200 cases) and the calibration team assessments (~200 cases). The new translations cut the mean motor scale difference in half, reducing it to around 2.4 points for the institutional team comparisons. The scoring difference in the motor scale was smaller for the calibration team comparisons—only one point—and interestingly, in contrast to the institutional teams, the FIM score was larger for the calibration teams. One reason for the

Table 5.1

Comparison of Motor and Cognitive Scales and Item-Level Means Across Instruments and Samples

	Institutional Team Assessments		Calibration Team Assessments	
FIM Scale or Item	FIM	PAC	FIM	PAC
Motor scale	45.46	47.73	46.80	45.76
Eating	5.51	5.54	5.73	5.77
Grooming	4.73	4.88	4.61	4.56
Bathing	3.24	3.30	3.27	3.03
Dressing upper body	4.25	4.35	3.92	3.95
Dressing lower body	2.99	3.21	2.80	2.77
Toileting	3.37	3.71	3.66	3.49
Bladder management	4.29	4.27	4.61	4.15
Bowel management	4.70	5.20	5.33	5.30
Transfer bed/chair	3.58	3.70	3.56	3.32
Transfer toilet	3.28	3.67	3.60	3.44
Transfer tub/shower	1.96	1.98	1.86	2.05
Locomotion—walk/wheelchair	2.22	2.20	2.51	2.38
Stairs	1.24	1.62	1.39	1.60
Cognitive scale	28.50	28.51	28.53	28.07
Comprehension	5.87	5.93	5.88	5.86
Expression	5.97	5.99	5.93	5.87
Social interaction	5.91	5.63	6.04	5.54
Problem solving	5.32	5.34	5.21	5.26
Memory	5.37	5.56	5.34	5.44

smaller difference may be that within the calibration team assessments, the same team is always completing both instruments for a specific case. The institutional assessments are a combination of cases, some of which are scored by the same teams, and some of which are not. Below, we examine how much difference this makes. All of the item-level means for both the institutional teams and the calibration teams were now within a half point of one another.

Next, we looked at the distributions of the motor and cognitive scales scored by institutional teams using the two instruments and at those scored by the calibration teams (see Table 5.2). The motor scale comparisons for the institutional teams show a clear pattern of increasing differences as one moves up the motor scale. At the low end, below 39, there is a one-point difference, which increases to four points for scores in the 60s and higher. There is no simple pattern to the differences for the calibration team motor scale assessments nor to either set of cognitive scale comparisons.

We use Pearson correlation coefficients (see Table 5.3) and kappa statistics (see Table 5.4) to help us quantify and understand the similarities and dissimilarities in scoring between these two instruments. In each case, the FIM score for an item

Table 5.2

Comparison of Motor and Cognitive Scale Distributions Across Instruments and Samples

	Institutional Team Assessments		Calibration Team Assessments	
Percentile	FIM	PAC	FIM	PAC
Motor scale				
1st	14	15	14	13
5th	23	23	21	20
10th	29	30	27	27
25th	38	39	37	38
50th	46	49	49	47
75th	54	57	57	55
90th	60	64	62	64
95th	63	67	65	67
99th	72	76	77	72
Cognitive scale				
1st	5	7	5	7
5th	12	13	10	10
10th	17	17	15	14
25th	25	25	25	25
50th	31	30	32	30
75th	35	35	35	35
90th	35	35	35	35
95th	35	35	35	35
99th	35	35	35	35

Table 5.3

Comparison of Pearson Correlation Coefficients Across Instruments and Samples

FIM Scale or Item	FIM vs. PAC: Institutional Team Assessments	FIM vs. PAC: Calibration Team Assessments
Motor scale	.85	.94
Eating	.73	.92
Grooming	.67	.74
Bathing	.61	.78
Dressing upper body	.74	.88
Dressing lower body	.71	.84
Toileting	.68	.84
Bladder management	.61	.76
Bowel management	.46	.70
Transfer bed/chair	.75	.80
Transfer toilet	.61	.85
Transfer tub/shower	.54	.74
Locomotion—walk/wheelchair	.57	.68
Stairs	.43	.53
Cognitive scale	.84	.96
Comprehension	.73	.83
Expression	.76	.87
Social interaction	.69	.80
Problem solving	.77	.89
Memory	.81	.91

is compared to the score on the pseudo-FIM from the MDS-PAC. We show these comparisons for both the institutional teams and the calibration teams. The correlations, which are more a measure of association than agreement, on the four scale comparisons (institutional team on motor scale, calibration team on motor scale, institutional team on cognitive scale, and calibration team on cognitive scale) generally meet the acceptable standard for scales (.85 or more). The calibration team correlations are substantially higher than those of the institutional teams (.94 vs. .85 on the motor scale and .96 vs. .84 on the cognitive scale). Item-level correlations are much lower. For the motor items scored by institutional teams, four of the 13 items were in the .4 to .6 range and the remainder were in the .6 to .8 range. The calibration team did better with only one motor item in the .4 to .6 range and six motor items in the .8 range and above. The correlations for the cognitive items are generally higher. For the institutional teams, four of the five items are in the .6 to .8 range and the fifth exceeds .8. For the calibration teams, all five items exceed .8.

A better measure of agreement is given by the kappa statistics (see Table 5.4) which measure agreement rather than association and correct for chance agreement. These indicate that the correspondence between the two instruments

Table 5.4

Comparison of Motor and Cognitive Scales and Item-Level Kappas Across Instruments and Samples

	FIM vs. PAC: Institutional Team Assessments		FIM vs. PAC: Calibration Team Assessments	
FIM Scale or Item	Simple Kappa	Weighted Kappa	Simple Kappa	Weighted Kappa
Motor scale				
Eating	.46	.61	.62	.81
Grooming	.43	.55	.53	.64
Bathing	.41	.53	.43	.60
Dressing upper body	.53	.64	.67	.80
Dressing lower body	.45	.59	.53	.71
Toileting	.37	.54	.45	.69
Bladder management	.22	.46	.38	.62
Bowel management	.17	.32	.50	.60
Transfer bed/chair	.42	.60	.44	.64
Transfer toilet	.37	.51	.48	.70
Transfer tub/shower	.43	.51	.64	.71
Locomotion—walk/wheelchair	.37	.49	.50	.61
Stairs	.31	.33	.53	.48
Cognitive scale				
Comprehension	.29	.54	.35	.65
Expression	.29	.56	.32	.65
Social interaction	.22	.48	.25	.55
Problem solving	.26	.57	.44	.73
Memory	.32	.62	.48	.77

is not as strong as it could be. For the institutional teams on both the motor and cognitive items, the weighted kappas are generally moderate. The simple kappas are often poor. This change in relative performance between the motor and cognitive items when we go from the correlations to the kappas results because the cognitive items all demonstrate a strong ceiling effect, with more than a quarter of all cases at the ceiling. Thus, when the kappa corrects for chance agreement (a relatively high probability of being at the ceiling), the true agreement is much lower. The calibration teams continue to do better. On the motor and cognitive items, the weighted kappas show good agreement with a few items in the moderate and a few in the very good range. Simple kappas for the calibration teams are higher than those of the institutional teams but not as high as the weighted kappas.

To help us understand more concretely what these statistics are telling us, we also provide data on the levels of absolute agreement and the amounts of disagreement (see Table 5.5). For the institutional teams and the bulk of the motor items, absolute agreement is between 50 to 70 percent. Three items—

Table 5.5

Comparison of Motor and Cognitive Scales and Item-Level Agreement Across Instruments and Samples

	FIM vs. PAC: Institutional Team Assessments				FIM vs. PAC: Calibration Team Assessments			
FIM Scale or Item	Agree	Within 1	Within 2	Off 2+	Agree	Within 1	Within 2	Off 2+
Motor scale	7.6	20.6	32.4	67.6	12.6	28.0	42.5	57.5
Eating	63.9	81.5	97.4	2.6	75.0	94.7	100.0	0.0
Grooming	63.1	82.9	96.7	3.3	74.0	88.9	97.1	2.9
Bathing	53.9	86.6	95.1	4.9	55.8	95.7	99.0	1.0
Dressing upper body	66.7	87.5	96.7	3.3	76.0	90.9	99.0	1.0
Dressing lower body	56.9	88.4	97.1	2.9	61.1	94.7	99.0	1.0
Toileting	48.6	83.0	93.7	6.3	53.9	91.8	98.6	1.4
Bladder management	35.0	62.3	80.5	19.5	51.0	72.6	87.0	13.0
Bowel management	31.8	60.0	77.3	22.8	64.9	81.7	88.0	12.0
Transfer bed	56.5	91.6	98.3	1.8	55.1	91.8	97.1	2.9
Transfer toilet	52.1	85.1	93.3	6.7	57.2	95.7	98.6	1.4
Transfer tub/shower	69.2	80.3	86.9	13.1	81.7	89.4	91.8	8.2
Locomotion—walk/ wheelchair	57.2	80.6	87.9	12.1	63.9	84.1	89.9	10.1
Stairs	78.1	87.2	89.2	10.8	83.7	89.9	89.9	10.1
Cognitive scale	23.6	40.5	55.5	44.5	32.2	60.0	71.7	28.3
Comprehension	51.1	87.7	97.2	2.8	55.8	90.9	99.5	0.5
Expression	53.1	88.3	97.6	2.4	54.3	90.9	100.0	0.0
Social interaction	44.2	80.6	94.3	5.7	47.6	81.1	96.1	3.9
Problem solving	42.6	79.9	95.8	4.2	56.3	89.4	98.6	1.4
Memory	48.8	85.3	96.4	3.6	60.9	93.7	98.6	1.4

bowel and bladder management and toileting—fall below this level. The highest level of agreement is on the stairs item, probably because of its large floor effect; many patients do not attempt stairs at admission and are consequently scored as total dependence on this item. For the calibration team, agreement on eight of the 13 motor items was between 50 and 70 percent. The remaining five items exceeded this level. On the five cognitive items, absolute agreement ranged from 43 to 53 percent for the institutional teams and from 48 to 61 percent for the calibration teams.

At the scale level, the picture on absolute agreement is somewhat different. For the motor scale, agreement was low, eight and 13 percent, for the institutional and calibration teams, respectively. For the cognitive scale, agreement was notably higher, 24 and 32 percent for the institutional and calibration teams, respectively. Although absolute agreement was not higher on the cognitive items relative to the motor items, the smaller number of items in the scale and the large ceiling effect work together to make the cognitive scale agreement level much higher than for the motor scale.

What Factors Contribute to These Observed Differences?

In Section 3, we identified some of the differences in scoring rules between the two instruments that could not be corrected in our translations. Here we try to discern the importance for scoring of the differences in the assessment periods on the two instruments. We also look at the role of others factors that could be contributing to these differences.

Differences in the Assessment Periods

As already noted, the MDS-PAC uses a three-day look-back period whereas the FIM looks only at the last 24 hours. Further, the FIM can be done anytime during the first 72 hours after admission. In theory, one might expect to see more of a scoring discrepancy for cases where the FIM is scored during the first 24 hours after admission and the MDS-PAC over the usual three-day assessment period as compared to cases where the FIM is scored on day three and the MDS-PAC over the three-day assessment period. If this were true, it could be part of why the institutional teams always had higher MDS-PAC scores. As part of our data collection effort, institutional data collectors were instructed to record the date of each assessment exam and the date on which they completed the FIM form. Just over a third of our institutional sample completed the FIM within the first 24 hours after admission. Just under a third completed the FIM on day three. Thus, with over a thousand cases in each group, we were able to analyze whether the timing of form completion was contributing to the scoring differences that we observed.

We were surprised to find that the timing did not seem to contribute to scoring differences. In Table 5.6, we show the means, standard deviations, and distribution of scores for both instruments when the FIM assessments are done within the first 24 hours in columns 2 and 3. The same data are shown for FIM assessments on day three in columns 4 and 5 and for all assessments in columns 6 and 7. The FIM scores are higher for the day three assessments, but so are the MDS-PAC scores. The mean difference actually increases from 2.41 to 2.55 as we go from day one to day three assessments and is lowest, 2.36, when we average across all the assessments. Table 5.7 presents the corresponding levels of absolute agreement and the correlations. Although the agreement on the motor score is about a point higher for day three assessments as compared to day one, the correlations are actually lower.

This relationship may be more complex than this analysis is able to reveal. Although roughly one-third of the cases in our sample appear to have been

Table 5.6

Comparison of Motor and Cognitive Scale Means and Distributions Across Instruments by FIM Assessment Day

	Day 1 Assessments		Day 3 Assessments		All Assessments	
	FIM	PAC	FIM	PAC	FIM	PAC
Motor scale percentile						
1	16	15	13	13	14	15
5	24	24	23	23	23	23
10	29	29	30	31	29	30
25	38	39	39	41	38	39
50	46	48	48	50	46	49
75	53	56	55	58	54	57
90	58	63	61	65	60	64
95	63	67	64	68	63	67
99	72	75	73	77	72	76
Mean	44.97	47.38	46.38	48.93	45.46	47.82
Standard deviation	11.62	12.94	12.18	13.25	12.01	13.14
Sample size	1,332		951		3,226	
Cognitive scale percentile						
1	6	7	5	7	5	7
5	12	13	12	13	12	13
10	17	17	17	17	17	17
25	25	24	25	25	25	25
50	30	30	31	32	31	30
75	34	35	35	35	35	35
90	35	35	35	35	35	35
95	35	35	35	35	35	35
99	35	35	35	35	35	35
Mean	28.04	28.33	28.65	28.70	28.50	28.51
Standard deviation	7.30	7.36	7.50	7.52	7.47	7.46
Sample size	1,303		1,146		3,310	

completed each day, the completion day is not random. Some facilities have "24 hour" policies that require staff to complete the FIM within the first 24 hours after admission. We tested whether the mean motor score differences were larger for those facilities than for all others and found that they were not. Next, we will explore these issues further using multivariate analysis.

Other Possible Contributors

Here we analyze motor score differences between the FIM and the MDS-PAC and the absolute value of motor score differences between the FIM and the MDS-PAC using regression analysis. Our model for D_{ij}, the MDS-PAC motor score – the FIM motor score, for patient i in facility j can be expressed as:

Table 5.7

MDS-PAC and FIM Agreement by FIM Assessment Day

FIM Scale or Item	Day 1 Assessments: Percentage in Absolute Agreement	Day 1 Assessments: Pearson Corre-lation	Day 3 Assessments: Percentage in Absolute Agreement	Day 3 Assessments: Pearson Corre-lation	All Assessments: Percentage in Absolute Agreement	All Assessments: Pearson Corre-lation
Motor scale	6.8	.86	8.3	.82	7.6	.85
Eating	63.1	.71	60.2	.73	63.9	.73
Grooming	65.4	.67	60.0	.67	63.1	.67
Bathing	55.0	.59	50.0	.59	53.9	.61
Dressing upper body	65.6	.70	65.5	.75	66.7	.74
Dressing lower body	57.2	.71	51.4	.66	56.9	.71
Toileting	48.9	.66	44.2	.66	48.6	.68
Bladder management	35.7	.59	35.4	.63	35.0	.61
Bowel management	26.8	.47	29.8	.50	31.8	.46
Transfer bed	55.8	.77	55.4	.71	56.5	.75
Transfer toilet	53.3	.63	50.3	.58	52.1	.61
Transfer tub/shower	70.4	.51	68.2	.58	69.2	.54
Locomotion—walk/ wheelchair	58.1	.62	54.2	.53	57.1	.57
Stairs	80.0	.44	74.7	.39	78.1	.43
Cognitive scale	18.9	.81	27.7	.85	23.6	.84
Comprehension	47.6	.70	53.8	.74	51.1	.73
Expression	49.5	.72	56.2	.79	53.1	.76
Social interaction	40.8	.66	47.3	.70	44.2	.69
Problem solving	37.6	.74	46.9	.79	42.6	.77
Memory	45.3	.79	50.0	.81	48.8	.81

$$D_{ij} = \alpha + X_{ij}\beta + Y_{ij}\delta + Z_j\gamma + \eta_j + \varepsilon_{ij}$$

The vector X_{ij} represents a set of variables that describe how instruments were administered, Y_{ij} represents the set of patient characteristics, and Z_j the set of hospital characteristics. We use a variance components model that includes a random effect, η_j, for hospitals. The variables on instrument administration include the timing (whether this case was in the first third of the study period, the middle segment, or the end), the number of people scoring the MDS-PAC (one person, two people, three people, or four or more people), whether the same or different teams scored the MDS-PAC and the FIM for this case, and the day on which the FIM motor score exam was completed. The set of patient characteristics includes patient age, gender, ability to communicate, and what RIC (1 = stroke, 7 = fracture of the lower extremity, 8 = replacement of lower extremity joint, any other) the patient is in. The set of hospital characteristics includes size (small, medium, and large), whether rural, and whether the hospital is freestanding (or hospital-based).

We began our analysis using ordinary least squares regression. We calculated studentized residuals and Cook's distance and looked at residual plots to identify outliers. We deleted one extreme outlier (motor score difference > 40) and all cases from the worst performing hospital. The mean difference in MDS-PAC and FIM motor scores for this facility was over 14 and we know from our training and certification procedures that this facility's team was the only team to fail the initial certification exam. When this hospital was included, it increased the variance resulting from the hospital effect by more than 50 percent. With these deletions, the residual plot showed no evidence of heteroscedasticity. Although the remaining studentized residuals did not pass a test for normality, the histogram and normal probability plots looked quite reasonable. We used PROC MIXED in SAS for the variance components estimation and looked at both intercept and slope effects (for the patient-level variables).

In the regression model, we found that team size was significant and that teams of size two or three had smaller differences than team of size one or four or more. The day on which the FIM motor exam was completed was also related to the scoring difference. When the motor part of the exam was completed on days one or two, the scoring differences were significant and larger than when scoring was done on day three. The coefficients on both the day one and the day two variables are positive, although the day two coefficient is larger and more statistically significant. In theory, we had expected day one to be the largest. Neither the period (beginning of the study, middle range, or late in the study) nor having different teams score the FIM and the MDS-PAC made a significant difference. Age, gender, and inability to communicate were not significant. We tested for differences among the larger RICs (stroke, lower extremity fracture, lower extremity joint replacement, and all other) and found the cases from RIC 8 were associated with larger scoring differences. With respect to hospital-level variables, none were significant; there was no difference between hospital-based and freestanding facilities nor between urban and rural, or by size. The random hospital effect was highly significant in both the intercept and slope for patients in RIC 8 and poor communication, indicating that hospitals differ systematically from one to the next and they differ in their effects on patients in RIC 8 and those with poor communication. The regression coefficients and p values are shown in Table 5.8.

Above, we analyzed motor score differences between the MDS-PAC and the FIM to determine whether there were variables that seemed to be associated with differences in scoring that led to systematic differences between the instruments. We also want to examine the absolute value of differences in scoring, that is, absolute scoring error. As before, we began with ordinary least squares regression. When we plotted the residuals, we observed that these were heteroscedastic with increasing variability for higher predicted absolute

Table 5.8

Regression Models for Motor Score Differences (PAC-FIM)

	Motor Score Differences		Ln (Σ│ PAC-FIM │)	
	Coefficient	p-Value	Coefficient	p-Value
Intercept	2.11	0.088	1.86	0.000
Period 1	–0.31	0.315	0.17	0.062
Period 2	–0.25	0.386	–0.14	0.098
Team size 1	–0.01	0.990	–0.02	0.923
Team size 2	–0.99	0.062	0.07	0.621
Team size 3	–1.08	0.011	–0.03	0.782
Different teams	0.43	0.229	0.22	0.036
FIM day 1	0.87	0.034	0.24	0.039
FIM day 2	1.11	0.004	0.11	0.325
Age < 65	–0.07	0.889	0.11	0.507
65 < age < 75	–0.03	0.944	–0.01	0.942
75 < age < 85	0.13	0.694	–0.01	0.952
Poor communication	–0.80	0.055	–0.95	0.000
Male	–0.26	0.293	0.07	0.330
RIC—joint replacement lower extremity	1.36	0.000	0.07	0.415
Small hospital	0.37	0.796	–0.18	0.493
Medium hospital	–2.27	0.064	–0.24	0.287
Rural hospital	–0.98	0.4854	0.02	0.926
Freestanding	–0.45	0.703	–0.15	0.492

differences. This led us to transform the dependent variable by taking its natural logarithm. When we examined the histogram of studentized residuals and the normal probability plot, these were improved under the transformation.

The results indicate that the use of different teams to score the MDS-PAC and the FIM significantly increased absolute scoring differences. Period and team size were unrelated to the absolute level of scoring differences. None of the hospital characteristics (size, rural, freestanding) were related to the level of absolute scoring differences. Similarly we observed no relationship between patient age or gender and the difference in scores. Poor patient ability to communicate reduced scoring difference significantly. The random effect for hospital was again highly significant although not as large as in the first model. In this model, the random slope effects we tested were not significant. Coefficients and p-values are given in columns 4 and 5 of Table 5.8.

The Role of Scoring Error

Although much of the difference in motor scores is inevitably due to differences in the scoring rules, some of it results from scoring error in both the MDS-PAC and the FIM. For example, some of the scoring rules are counterintuitive and this may lead to increased scoring error. The MDS-PAC's ADL Assist Codes are an area where this may happen. The instructions are to score one person doing weight-bearing assistance with one limb assist as 1, but to combine all other one-person assists (torso, multiple limb, and non-weight-bearing) in with the no assistance group. Another example occurs in the scoring of bowel and bladder appliance support when medications are used. When a nurse hands an otherwise independently functioning patient a medication, coders were instructed to score the MDS-PAC's appliance support item as maximal assistance. Since hospital policy requires that the nurse hand the patient the medication unless patients are on special medication programs (which is unlikely at admission), these patients are all scored as highly dependent on the MDS-PAC. The situation is exacerbated by including Metamucil®, a fiber enhancement given to most hospitalized patients, on the medication list.

The simplicity of the FIM scoring sheets is deceptive, as the actual scoring rules are quite complicated. Here, we show two examples where explicit questions on the MDS-PAC allow us to observe how FIM scorers have overlooked some of the scoring nuances. In the eating item, FIM scoring rules instruct that patients with chewing problems and those on modified diets for swallowing problems who would otherwise be scored as independent in eating, should be scored as modified independence. Since the MDS-PAC explicitly records both chewing problems and diet modifications for swallowing problems, we were able to observe this. Assuming that the PAC items are measured without error, we infer that these patients were scored correctly only 43 percent of the time. Fortunately, chewing problems and/or diet modifications occur in only 6 percent of the otherwise independent eaters. Another example comes in the FIM locomotion item, which has a distance component. FIM scoring rules clearly require (with one exception) that patients go at least 150 feet to score above maximum assistance. Since the MDS-PAC includes an explicit item on distance, if we are again willing to assume that the explicit distance response is correct (or measured with low error), we observe that FIM scorers overlooked the distance requirements 16 percent of the time. Among patients who were unable to walk 150 feet, 24 percent were scored incorrectly. By including these elements explicitly as part of the overall item, one may encourage and improve scoring accuracy.

Summary

In this section, we used the new translation for motor items and demonstrated that it did, in fact, reduce the mean difference in motor scores between the FIM and the MDS-PAC by 50 percent. Despite the improvement, we find that the agreement between the instruments as measured by kappa statistics is poor to moderate. Comparing the levels of instrument agreement between the institutional teams and our calibration teams, we find that the latter have notably higher levels of agreement. We discuss some of the differences in scoring rules between the instruments that preclude perfect agreement. As anticipated, differences in the assessment periods between the instruments contributed to the mean difference in motor scores. Other factors that influenced scoring differences were the size of the team scoring the MDS-PAC and whether the patient was in RIC 8, lower extremity joint replacement. After controlling explicitly for the variables that we could, we found that a hospital-specific intercept and slope effects remained. Finally, we provide some information on scoring errors and speculate on how these could be reduced as another strategy for reducing the motor score differences.

6. Mapping Pseudo-FIM Motor and Cognitive Scores into CMGs

Introduction

In this section, we consider several methods of transforming the pseudo-FIM items and scales (from the translation described in Section 3) to reduce the remaining differences with the actual FIM item and scale scores. To clarify the process, we first provide background information on the structure of CMGs and how they were created. Then, we describe the different transformations that we considered. Next, we use the transformed motor and cognitive scores and compare how often these map into the same case mix group that the actual FIM scores would have mapped into. Since each case mix group has an associated payment weight, we then look at the differences in payment that result from classifying cases using the MDS-PAC with its pseudo-FIM scores compared to those obtained directly from the FIM. We compare these payments both overall and at the facility level. Finally, we look for any systematic differences by type of patient or type of facility.

CMGs

CMGs have their origin in a patient classification system called Functional Independence Measure- Function Related Groups (FIM-FRGs) developed by Stineman et al. (1994a). Using the basic structure of FIM-FRGs together with newer and much larger datasets, RAND has updated, refined, and completed this classification work making it a suitable foundation for the new payment system (Carter et al., 2002a).

In this classification system, rehabilitation providers indicate the reason the patient is in inpatient therapy by selecting the appropriate impairment group. These impairment groups aggregate up to a Rehabilitation Impairment Category using assignment rules developed by RAND.

Classification and regression tree (CART) analysis was used on the set of patients within each RIC to define the actual payment cells for that RIC. Three variables were used in the CART analysis, the FIM motor score (range 13–91), the FIM cognitive score (range 5–35), and patient age. So, for example, in the version of

CMGs that we used, CART divided the stroke RIC into 11 payment cells as follows:

Cell 1	Motor score between 62 and 91
Cell 2	Motor score between 57 and 61; cognitive score between 27 and 35
Cell 3	Motor score between 51 and 56; cognitive score between 28 and 35
Cell 4	Motor score between 57 and 61; cognitive score between 5 and 26
Cell 5	Motor score between 51 and 56; cognitive score between 2 and 27
Cell 6	Motor score between 46 and 50
Cell 7	Motor score between 42 and 45
Cell 8	Motor score between 36 and 41
Cell 9	Motor score between 13 and 35; age $\geq$ 84
Cell 10	Motor score between 31 and 35; age $\leq$ 83
Cell 11	Motor score between 13 and 30; age $\leq$ 83

In the work described in the next subsections, we use several methods to transform the pseudo-FIM motor and cognitive scores so that they will better match the FIM motor and cognitive scores. We use the pseudo-FIM motor and cognitive scores (raw and transformed) together with patient age to map patients into payment cells as described above. For the analysis of agreement, we compare whether a given patient is mapped into the same payment cell when his pseudo-FIM motor and cognitive scores are used as he would have been had the actual FIM scores been used. To illustrate, suppose a patient has a motor score of 44 on the FIM and 46 on the PAC's pseudo-FIM, then he would map into Cell 7 using the FIM and Cell 6 using the pseudo-FIM.

Mapping and Adjusting Pseudo-FIM Scores to Match FIM Scores

Our most basic approach to mapping the MDS-PAC data into case mix groups is to take the 13 pseudo-FIM motor items created using the translation in Section 3 and sum them to form an MDS-PAC motor scale. Similarly, we can take the five pseudo-FIM cognitive items and sum them to form an MDS-PAC cognitive scale. Using the patient's RIC and age, together with the MDS-PAC motor and MDS-PAC cognitive scale scores from the MDS-PAC, we can now classify patients into case mix groups as described above. We then compare how often these agree. We refer to this as our unadjusted or baseline case.

Since the motor scale correlation was fairly high, we reasoned that an approach that standardized the MDS-PAC motor elements to the FIM elements could improve the agreement. To operationalize this, we multiplied each MDS-PAC pseudo-FIM element by the ratio of the standard deviation of the FIM element to the standard deviation of the corresponding MDS-PAC pseudo-FIM element and then subtracted the difference between the corresponding element means, that is, the bias. After standardizing each individual element, we computed a new motor scale score by summing the standardized elements and computing the standard deviation of this new summed scale. The last step was to standardize the sum by multiplying by the ratio of the standard deviation of the FIM motor scale to the standard deviation of the sum of the standardized elements and then subtracting any remaining difference between the mean of the sum of the standardized elements and the mean of the FIM motor scale. The ratio of the standard deviations for each element and the whole scale is shown in column 2 of Table 6.1. The mean differences, which are then subtracted, are shown in column 3. Unlike the previous examples, in this case we made these adjustments to the cognitive scale as well.

Finally, we used a regression approach to model the actual FIM motor score as a linear combination of the 13 MDS-PAC pseudo-FIM motor items. Similarly, we modeled the actual FIM cognitive score as a linear combination of the five MDS-PAC pseudo-FIM cognitive items. The regression coefficients for these models are shown in Table 6.2.

Since our ultimate goal is not to map one scale into the other but rather to look at whether and how these mappings influence agreement and disagreement when cases are classified into case mix groups, we also want to consider the effects of sample refinement on this agreement. Using the regression approach, we calculate how CMG agreement changes when the hospital with the largest disagreement on motor scores is excluded. We also consider how restricting the sample to cases for which the MDS-PAC and the FIM were scored by the same team influences agreement on CMG classification. Because we corrected the impairment group codes for 6 to 9 percent of the cases, we also wanted to make sure that these had not adversely affected agreement, so we calculated the level of agreement without these cases as well. Our last analysis on agreement restricts the sample to the calibration team MDS-PAC and FIM cases and provides information on level of agreement with calibration team scoring.

Table 6.1

Adjustments to Standardize Pseudo-FIM Motor and Cognitive Scales to Actual FIM Motor and Cognitive Scale Means and Standard Deviations

FIM Scale or Item	σ_{FIM} / σ_{PAC}	μ_{PAC} - μ_{FIM}
Motor scale	0.94	0.011
Eating	1.03	0.196
Grooming	0.96	-0.039
Bathing	0.97	-0.046
Dressing upper body	1.01	0.140
Dressing lower body	1.02	0.270
Toileting	1.01	0.379
Bladder management	0.93	–0.338
Bowel management	0.87	–0.155
Transfer bed/chair	0.94	–0.104
Transfer toilet	1.06	0.603
Transfer tub/shower	0.94	–0.101
Locomotion—walk/wheelchair	0.94	–0.161
Stairs	0.59	–0.278
Cognitive scale	0.96	0.003
Comprehension	1.09	0.619
Expression	1.17	1.019
Social interaction	0.94	–0.607
Problem solving	1.01	0.084
Memory	1.05	0.463

Accuracy of Alternative Mappings

In Table 6.3, we report on the agreement and disagreement for each of these cases. We restrict our sample to the set of cases for which FIM and MDS-PAC scorers placed patients into the same RIC. Had we not done this, the level of agreement would have been 2–5 percent lower. We find that for our base case, which includes all cases with matching RICs and no motor score adjustment, in 53 percent of the cases, the MDS-PAC and the FIM map into the same CGM. We also used an ad hoc adjustment that attempted to align the percentiles of the MDS-PAC and FIM motor distributions (Table 6.2, columns 2 and 3) as follows: (1) for MDS-PAC motor scale values between 15 and 39, deduct 1 point; (2) for MDS-PAC motor scale values between 40 and 45, deduct 2 points; (3) for MDS-PAC motor scale values between 46 and 59, deduct 3 points; and (4) for MDS-PAC motor scale values of 60 or more, deduct 4 points. This did very little to improve the accuracy and is not included in the table.

The standardization approach increases agreement by about 2 percentage points and the regression approach increases it another point. Deleting the facility with the worst motor score agreement increases agreement to 57 percent. Other

Table 6.2

Regression Coefficients for Transforming Pseudo-FIM Motor and Cognitive Items to Actual FIM Motor and Cognitive Scales

Pseudo-FIM Item	Regression Coefficients	p-Values
FIM Motor Scale		
Intercept	7.34	0.0001
Eating	0.99	0.0001
Grooming	0.73	0.0001
Bathing	0.70	0.0001
Dressing upper body	1.31	0.0001
Dressing lower body	1.09	0.0001
Toileting	1.04	0.0001
Bladder management	0.68	0.0001
Bowel management	0.35	0.0001
Transfer bed/chair	1.28	0.0001
Transfer toilet	0.47	0.0083
Transfer tub/shower	0.45	0.0001
Locomotion—walk/wheelchair	0.73	0.0001
Stairs	–0.07	0.4569
R**2	0.75	
Sample size	3,225	
FIM Cognitive Scale		
Intercept	5.34	0.0001
Comprehension	–0.78	0.0019
Expression	1.78	0.0001
Social interaction	0.97	0.0001
Problem solving	0.94	0.0001
Memory	1.19	0.0001
R**2	0.71	
Sample size	3,310	

Table 6.3

Agreement and Disagreement on Case Mix Group Classification Within RICs Between the FIM and the MDS-PAC

Mapping/Sample	Agreed	Within 1	Within 2	Within 5	> 5
No adjustment	.53	.84	.93	0.999	.002
Standardization	.55	.85	.94	0.999	.001
Regression—all cases	.56	.86	.94	0.999	.001
Regression deleting worst	.57	.86	.95	0.999	.001
Regression—calibration team cases	.67	.90	.97	1.000	0.0

adjustments and sample restrictions had even smaller effects. In Section 5, we saw that when the calibration teams scored both the MDS-PAC and FIM, that motor score agreement was much closer. Considering only these cases, we find that CMG agreement reaches 67 percent, notably better than what we were able to achieve with any of our adjustments.

We looked at agreement by RIC and found that it ranged from 49 percent for RIC 20, miscellaneous, up to 83 percent for RIC 18, multiple major trauma with brain or spinal cord injury (see Table 6.4). Nine of the RICs had agreement in the 60–69 percent range, but agreement tended to be good only in small RICs and was worst among the three largest RICs (stroke, lower extremity joint replacement, and miscellaneous). The top five RICs in terms of agreement constituted only 6 percent of the sample compared to 61 percent in the worst five.

To provide some perspective on this level of agreement when using two instruments, we calculated the CMG agreement achieved when different scoring teams used the same instrument. We did this comparison between the institutional teams and the calibration teams for both the FIM and the MDS-PAC. We began by requiring that the RICs match before calculating the CMG agreement, as we had done for comparisons of the institutional teams' MDS-PAC and FIM classifications. Next, we looked at the classification agreement between raters on the same instrument (50 to 55 percent) and found that it is in the same range as the between-instrument agreement (53 to 56 percent). These low levels of inter-rater agreement in classification are quite troubling (see Table 6.5). If within-instrument scoring error is this large, then it cannot be surprising that between-instrument agreement is low.

Payment Differences

Across the 2,959 cases in our sample, the mean per case difference in payments between the MDS-PAC and the FIM is surprisingly small, only –$46 and not significantly different from zero. The range of payment differences (PAC – FIM) is large going from a minimum of –$14,227 to +$9,765 and the tails are quite thick. As the agreement data that we just reviewed indicate, payments are equal for 50 to 60 percent of the cases. However, for more than 20 percent of the individual cases, the payment differences exceed $2,000 in one direction or the other. With such large payment differences, we will necessarily be concerned about whether these errors tend to cluster and about what happens at the facility level.

The mean per case payment difference for the 50 facilities in our sample was $82 with a standard deviation of $837; that is, cases classified using the MDS-PAC on average received $82 more than when the same cases were classified using the FIM. The distribution of per case differences was actually normally distributed and the mean was not significantly different from zero. For 30 of the 50 facilities,

Table 6.4

Agreement at the RIC and CMG Level

	Absolute Agreement	Within 1 CMG	Within 2 CMGs	Number of Cases
All cases	56.4	86.1	94.5	2,959
RIC 1: stroke				
All	50.4	76.3	87.4	581
CMG 1	43.9			41
CMG 2	18.8			16
CMG 3	35.3			34
CMG 4	22.2			9
CMG 5	40.5			37
CMG 6	40.8			76
CMG 7	32.4			74
CMG 8	45.6			79
CMG 9	85.7			49
CMG 10	40.8			49
CMG 11	76.9			117
RIC 2: traumatic brain injury				
All	57.9	86.8	100.0	38
CMG 1	—			2
CMG 2	40.0			5
CMG 3	91.7			12
CMG 4	33.3			3
CMG 5	50.0			8
CMG 6	50.0			8
RIC 3: non-traumatic brain injury				
All	63.1	87.7	98.5	65
CMG 1	54.6			11
CMG 2	—			1
CMG 3	71.4			21
CMG 4	53.3			15
CMG 5	70.6			17
RIC 4: traumatic spinal cord				
All	61.9	100.0	100.0	21
CMG 1	0			1
CMG 2	81.8			11
CMG 3	80.0			5
CMG 4	—			4
RIC 5: non-traumatic spinal cord				
All	59.0	99.0	100.0	100
CMG 1	—			0
CMG 2	54.6			22
CMG 3	51.6			31
CMG 4	63.0			27
CMG 5	70.0			20
RIC 6: neurological				
All	68.1	95.1	100.0	163
CMG 1	58.3			24

Table 6.4 (continued)

	Absolute Agreement	Within 1 CMG	Within 2 CMGs	Number of Cases
CMG 2	72.2			54
CMG 3	57.6			33
CMG 4	75.0			52
RIC 7: fracture of lower extremity				
All	60.2	94.1	100.0	271
CMG 1	50.0			26
CMG 2	60.9			87
CMG 3	40.9			71
CMG 4	78.2			87
RIC 8: replacement of lower extremity joint				
All	51.7	88.0	97.0	755
CMG 1	42.1			88
CMG 2	62.0			300
CMG 3	47.5			223
CMG 4	32.5			40
CMG 5	36.2			69
CMG 6	65.7			35
RIC 9: other orthopedic				
All	64.7	100.0	100.0	102
CMG 1	40.0			10
CMG 2	72.5			40
CMG 3	59.4			32
CMG 4	70.0			20
RIC 10: amputation, lower extremity				
All	71.8	100.0	100.0	78
CMG 1	61.5			13
CMG 2	60.9			23
CMG 3	81.0			42
RIC 11: amputation, other				
All	66.7	83.3	100.0	6
CMG 1	66.7			3
CMG 2	—			0
CMG 3	100.0			1
CMG 4	50.0			2
RIC 12: osteoarthritis				
All	66.2	97.2	100.0	71
CMG 1	66.7			21
CMG 2	66.7			18
CMG 3	63.6			22
CMG 4	70.0			10
RIC 13: rheumatoid, other arthritis				
All	80.0	100.0	100.0	15
CMG 1	—			3
CMG 2	100.0			8
CMG 3	100.0			4

Table 6.4 (continued)

	Absolute Agreement	Within 1 CMG	Within 2 CMGs	Number of Cases
RIC 14: cardiac				
All	67.2	100.0	100.0	134
CMG 1	70.0			50
CMG 2	69.8			63
CMG 3	52.4			21
RIC 15: pulmonary				
All	71.7	87.0	100.0	46
CMG 1	33.3			6
CMG 2	90.0			20
CMG 3	70.6			17
CMG 4	33.3			3
RIC 16: pain syndrome				
All	69.1	95.2	100.0	42
CMG 1	58.3			12
CMG 2	50.0			12
CMG 3	88.9			18
RIC 17: MMT-NBSCI				
All	79.2	100.0	100.0	24
CMG 1	75.0			12
CMG 2	83.3			12
RIC 18: MMT-BSCI				
All	83.3	100.0	100.0	6
CMG 1	100.0			2
CMG 2	75.0			4
RIC 19: Guillain-Barre				
All	57.1	100.0	100.0	7
CMG 1	50.0			2
CMG 2	50.0			2
CMG 3	66.7			3
RIC 20: miscellaneous				
All	49.1	69.4	84.8	434
CMG 1	20.0			5
CMG 2	11.9			42
CMG 3	30.2			63
CMG 4	—			0
CMG 5	50.6			89
CMG 6	61.6			112
CMG 7	54.6			11
CMG 8	82.6			23
CMG 9	51.3			39
CMG 10	43.8			16
CMG 11	64.7			34

the mean per case MDS-PAC payment exceeded the FIM payment. Although these numbers are somewhat more reassuring than the level of classification error led us to anticipate, it is important to note that for more than 20 percent of the facilities, the mean per case difference exceeded $1,000 in one direction or the other, ranging

Table 6.5

Within-Instrument Agreement and Disagreement on Case Mix Group Classification Within RICs

	Agreed	Within 1	Within 2	Within 5	> 5
Institutional FIM to calibration FIM	.50	.88	.93	1.00	0.0
Institutional MDS-PAC to calibration MDS-PAC	.55	.84	.93	1.00	0.0

from –$2,434 to +$2,438. For these facilities, the difference in revenue between the two instruments is 10 to 20 percent in one direction or the other. The magnitude of these differences and the number of facilities affected is of concern. The number of cases per facility ranged from a low of three to a maximum of 154. If we limit the analysis to hospitals with more than 50 cases, then the mean across facilities is –$155; dropping the worst performing hospital from this group reduces the mean to –$67, but we are still left with around 20 percent of the hospitals in this group with mean per case differences in excess of $1,000.

When we looked at payment differences across RICs, we observed large negative differences for three of the small RICs, traumatic spinal cord injury (–$2,710 across 21 cases), multiple major trauma with brain or spinal cord injury (–$1,965 across six cases), and Guillain-Barre (–$1,547 across seven cases). Differences on the three big RICs were not large; $127 for stroke, –$44 for lower extremity fracture, and –$153 for lower extremity joint replacement. The miscellaneous group had a somewhat larger difference, $240. The remaining differences ranged from –$549 to + $273 (seeTable 6.6).

Regression Analysis of Payment Differences

In this subsection, we use regression analysis to examine whether the observed payment differences are associated with certain types of patients or types of hospitals. We look for systematic differences that are influencing payments more in one direction or the other and we look for characteristics associated with high scoring error in either direction. The model we use is:

$$P_{ij} = \alpha + Y_{ij}\delta + Z_j\gamma + \eta_j + \varepsilon_{ij}$$

where P_{ij} denotes the MDS-PAC payment – the FIM payment for patient i in hospital j, Y_{ij} is a vector of patient characteristics, and Z_j is a vector of hospital characteristics. The patient characteristics that we test include age, marital status, education, race, gender, and poor ability to communicate. The hospital characteristics include whether the facility is freestanding or hospital-based, its

Table 6.6

Means and Standard Deviations of Payment Levels

	Number of Cases	MDS-PAC Payment	FIM Payment	PAC-FIM
Overall	2,959	11,142	11,188	–46
		(3,978)	(4,219)	(1,960)
By hospital	50			82
				(837)
By hospitalfor hospital > 50 cases	27			–155
				(733)
By RIC				
1 stroke	581	15,292	15,164	127
		(4,376)	(4,731)	(2,227)
2 traumatic brain injury	38	14,072	14,600	–528
		(4,413)	(5,906)	(3,468)
3 non-traumatic brain injury	65	13,071	13,351	–281
		(4,089)	(4,565)	(2,783)
4 traumatic spinal cord injury	21	15,053	17,763	–2,710
		(4,269)	(8,823)	(6,368)
5 non-traumatic spinal cord injury	100	11,634	11,943	–309
		(4,417)	(4,744)	(2,756)
6 neurological	163	11,674	11,846	–172
		(2,380)	(2,603)	(1,602)
7 lower extremity fracture	271	10,848	10,892	–44
		(1,969)	(1,964)	(1,505)
8 lower extremity joint replacement	755	7,711	7,864	–153
		(1,803)	(1,915)	(1,459)
9 other orthopedic	102	10,042	10,057	–15
		(2,072)	(2,265)	(1,459)
10 amputation, lower extremity	78	13,720	13,718	1
		(2,187)	(2,295)	(1,618)
11 amputation, other	6	10,188	10,737	–549
		(3,298)	(4,206)	(2,772)
12 osteoarthritis	71	9,083	9,005	78
		(2,074)	(2,138)	(1,322)
13 rheumatoid, other arthritis	15	8,993	8,720	273
		(1,726)	(1,972)	(566)
14 cardiac	134	9,758	9,845	–87
		(2,277)	(2,440)	(2,004)
15 pulmonary	46	11,494	11,919	–424
		(2,936)	(4,209)	(3,370)
16 pain syndrome	42	9,213	8,954	259
		(1,770)	(1,776)	(1,380)
17 MMT–NBSCI	24	12,810	12,561	249
		(3,041)	(3,052)	(2,774)
18 MMT–BSCI	6	16,912	18,877	–1,965
		(6,456)	(6,087)	(4,812)
19 Guillain-Barre	7	13,949	15,497	–1,547
		(6,404)	(7,140)	(4,773)
20 miscellaneous	434	11,273	11,032	240
		(2,400)	(2,688)	(1,467)

size, and whether it is in an urban or rural location. Our model includes a random effect, $\eta_{j,}$ for hospitals and we test whether the random hospital effect has slope components or only an intercept effect.

We find that none of the patient or hospital characteristics included in our model are associated with systematic differences in payment, a positive finding. However, the random effect for hospitals is again highly significant, indicating significant scoring differences between hospitals that are not captured by any of these variables.

In addition to looking at the real difference in payment, we wanted to model the absolute value of the differences to see if we could identify characteristics associated with scoring error. To do this, we use a two-part model. First, we model the probability of having a non-zero payment difference using logistic regression. Then, we model the natural log of the absolute value of that difference.

Our logistic regression model of the probability of non-zero payment difference indicates that nonwhites, those with poor ability to communicate, and patients from rural hospitals are more likely to have some form of payment difference. For nonwhites and those from rural hospitals the odds ratios are 1.4; for those with poor ability to communicate the odds ratio increases to 1.8. Our model did not include a random hospital effect. With a random hospital effect, some of these variables may no longer be significant.

The second part of this model estimates the log of the absolute value of non-zero payment differences. Greater absolute differences were significant for males and those with poor ability to communicate. The random intercept effect for hospitals was significant; the random slope effects we tested were not. This indicates that hospitals will differ systematically in their payment levels but these differences are poorly explained by the observable patient and hospital characteristics that we were able to test (see Table 6.7).

Summary

In this section, we mapped FIM and pseudo-FIM motor and cognitive scores into CMGs and tried several approaches to improve the match between the mappings. We found that the FIM and the MDS-PAC mapped into the same CMG 53 percent of the time initially. We improved the level of agreement to 57 percent by using a regression mapping of pseudo-FIM items onto the FIM scores and by dropping the worst facility.

Table 6.7

Regression Models on Payment Differences and Absolute Payment Differences

	PAC – FIM Payment	P(Absolute (PAC – FIM Payment) > 0)	Ln(Absolute (PAC – FIM Payment))
Intercept	32.6	–0.02	7.64***
Age < 65	–115.6	0.10	0.02
65 < age < 75	81.0	—	0.03
75 < age < 85	–19.5	–0.04	0.06
Age > 85	—	0.21	—
Married	101.3	0.06	–0.06
No high school	41.6	–0.04	–0.03
Some college	2.7	0.12	–0.06
Nonwhite	178.8	0.31**	–0.03
Poor communication	–102.9	0.60***	0.25***
Male	61.4	–0.04	0.11***
Small hospital	–468.0	—	0.08
Medium hospital	225.7	0.06	0.02
Large hospital	—	0.13	—
Rural hospital	425.4	0.36**	–0.01
Freestanding	–169.9	0.07	–0.01

*p < .05; **p < .01; ***p < .001.

CMG agreement within RICs was best for a few small RICs, although it was generally much lower among the larger RICs. Although this level of agreement between instruments appears low for use in a payment system, we found that scoring error within instruments was high and led to equally poor agreement.

Despite the poor levels of classification agreement, mean payment differences between the two instruments were small, averaging –$46, and not significantly different from zero. At the facility level, mean per case differences increased somewhat to $82. These differences are not large, but we are concerned because more than 20 percent of the facilities would experience revenue differences of 10 percent or more. Our multivariate analysis of payment differences showed significant differences across hospitals but these were not systematically associated with patient or hospital characteristics. Race, gender, poor ability to communicate, and coming from a rural hospital increased the likelihood of payment differences and/or the absolute amount of the difference.

References

Carter, G. M., J. L. Buchanan, T. Donyo, M. Inkelas, and K. L. Spritzer, *A Prospective Payment System for Inpatient Rehabilitation,* National Technical Information Service, publication number PB98106024, available at www.ntis.gov, Springfield, VA, 1997a.

Carter, G. M., D. A. Relles, and J. L. Buchanan, *A Classification System for Inpatient Rehabilitation Patients: A Review and Proposed Revisions to the Functional Independence Measure–Function Related Groups,* National Technical Information Service, publication number PB98105992, available at www.ntis.gov, Springfield, VA, 1997b.

Carter, G. M., J. L. Buchanan, M. B. Buntin, et al., *Executive Summary of Analyses for the Initial Implementation of the Inpatient Rehabilitation Facility Prospective Payment System*, Santa Monica, CA: RAND, MR-1500/1-CMS, 2002a.

Carter, G. M., M. B. Buntin, J. Hayden, et al., *Analyses for the Initial Implementation of the Inpatient Rehabilitation Facility Prospective Payment System,* Santa Monica, CA: RAND, MR-1500-CMS, 2002b.

Carter, G. M., D. A. Relles, B. O. Wynn, et al., *Interim Report on an Inpatient Rehabilitation Facility Prospective Payment System,* Santa Monica, CA: RAND, MR-1503-CMS, 2002c.

Casten, R., M. P. Lawton, P. A. Parmelee, and M. H. Kleban, "Psychometric Characteristics of the Minimum Data Set I: Confirmatory Factor Analysis," *J Amer Ger Soc,* Vol. 46, 1998, pp. 726–735.

Centers for Medicare and Medicaid Services, "Medicare Program; Prospective Payment System for Inpatient Rehabilitation Facilities. Final Rule," *Federal Register*, Vol. 66, No. 152, August 7, 2001.

Corrigan, J. D., K. Smith-Knapp, and C. V. Granger, "Validity of the Functional Independence Measure for Persons with Traumatic Brain Injury," *Arch Phys Med Rehabil,* Vol. 78, No. 8, 1997, pp. 828–834.

Dodds, T. A., D. P. Martin, W. C. Stolov, and R. A. Deyo, "A Validation of the Functional Independence Measurement and Its Performance Among Rehabilitation Patients," *Arch Phys Med Rehabil*, Vol. 74, 1993, pp. 531–536.

Eilertsen, T. B., A. M. Kramer, R. E. Schlenker, and C. A. Hrincevich, "Application of Functional Independence Measure–Function Related Groups and Resource Utilization Groups—Version III Systems Across Post-Acute Care Settings," *Med Care*, Vol. 36, No. 5, 1998, pp. 695–705.

Fleiss, J. L., *Statistical Methods for Rates and Proportions,* 2nd edition, New York: John Wiley, 1981.

Fries, B. E., D. P. Schneider, W. J. Foley, M. Gavazzi, R. Burke, and E. Cornelius, "Refining a Case-Mix Measure for Nursing Homes: Resource Utilization Groups (RUG III)," *Med Care*, Vol. 32, No. 7, 1994, pp. 668–685.

Hamilton, B. B., J. A. Laughlin, R. C. Fiedler, and C. V. Granger, "Inter-rater Reliability of the 7-Level Functional Independence Measure (FIM)," *Scan J Rehab Med*, Vol. 26, 1994, pp. 115–119.

Hartmaier, S. L., P. D. Sloane, H. A. Guess, and G. G. Koch, "The MDS Cognition Scale: A Valid Instrument for Identifying and Staging Nursing Home Residents with Dementia Using the Minimum Data Set," *J Amer Ger Soc*, Vol. 42, 1994, pp. 1173–1179.

Hartmaier, S. L., P. D. Sloane, H. A. Guess, G. G. Koch, C. M. Mitchell, and C. D. Phillips, "Validation of the Minimum Data Set Cognitive Performance Scale: Agreement with the Mini-Mental State Examination," *J Ger: Med Sci*, Vol. 50A, No. 2, 1995, pp. M128–M133.

Hawes, C., J. N. Morris, C. D. Phillips, V. Mor, B. E. Fries, and S. Nonemaker, "Reliability Estimates for the Minimum Data Set for Nursing Home Resident Assessment and Care Screening (MDS)," *The Gerontologist*, Vol. 35, No. 2, 1995, pp. 171–178.

Hawes, C., J. N. Morris, C. D. Phillips, B. E. Fries, K. Murphy, and V. Mor, "Development of the Resident Assessment Instrument in the USA," *Age and Aging*, Vol. 26-S2, 1997, pp. 19–25.

Health Care Financing Administration, "Medicare Program; Prospective Payment System for Inpatient Rehabilitation Facilities," *Federal Register*, Notice of Proposed Rule Making, November 3, 2000.

Heinemann, A. W., J. M. Linacre, B. D. Wright, B. B. Hamilton, and C. V. Granger, "Relationships Between Impairment and Physical Disability as Measured by the Functional Independence Measure," *Arch Phys Med Rehabil*, Vol. 74, 1993, pp. 566–573.

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

Kramer, A. M., J. F. Steiner, R. E. Schlenker, T. B. Eilertsen, C. A. Hrincevich, D. A. Tropea, L. A. Ahmad, and D. G. Eckhoff, "Outcomes and Costs After Hip Fracture and Stroke: A Comparison of Rehabilitation Settings," *JAMA*, Vol. 277, No. 5, 1997, pp. 396–404.

Lawton, M. P., R. Casten, P. A. Parmelee, K. Van Haitsma, J. Corn, and M. H. Kleban, "Psychometric Characteristics of the Minimum Data Set II: Validity," *J Amer Ger Soc*, Vol. 46, 1998, pp. 736–744.

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

Morris, J. N., C. Hawes, B. E. Fries, C. D. Phillips, V. Mor, S. Katz, K. Murphy, M. L. Drugovich, and A. S. Friedlob, "Designing the National Resident

Assessment Instrument for Nursing Homes," *The Gerontologist*, Vol. 30, No. 3, 1990, pp. 293–307.

Morris, J. N., B. E. Fries, D. R. Mehr, C. Hawes, C. Phillips, V. Mor, and L. A. Lipsitz, "MDS Cognitive Performance Scale," *J Ger: Med Sci*, Vol. 49, No. 4, 1994, pp. M174–M182.

Ottenbacher, K. J., Y. Hsu, C. V. Granger, and R. C. Fiedler, "The Reliability of the Functional Independence Measure: A Quantitative Review," *Arch Phys Med Rehabil*, Vol. 77, 1996, pp. 1226–1232.

Pollack, N., W. Rheault, and J. L. Stoecker, "Reliability and Validity of the FIM for Persons Aged 80 Years and Above from a Multilevel Continuing Care Retirement Community," *Arch Phys Med Rehabil*, Vol. 77, No. 10, 1996, pp. 1056–1061.

Ravaud, J. F., M. Delcey, and A. Yelnik, "Construct Validity of the Functional Independence Measure (FIM): Questioning the Unidimensionality of the Scale and the Value of FIM Scores," *Scand J Rehab Med*, Vol. 31, 1999, pp. 31–41.

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

Sgadari A., J. N. Morris, B. E. Fries, G. Ljunggren, P. V. Jonsson, J. N. DuPaquier, and M. Schroll, "Efforts to Establish the Reliability of the Resident Assessment Instrument," *Age and Aging*, Vol. 26-S2, 1997, pp. 27–30.

Sharrack B., R.A.C. Hughes, S. Soudain, and G. Dunn, "The Psychometric Properties of Clinical Rating Scales Used in Multiple Sclerosis," *Brain*, Vol. 122, 1999, pp. 141–159.

Snowden M., W. McCormick, J. Russo, D. Srebnik, K. Comtois, J. Bowen, L. Terri, and E. B. Larson, "Validity and Responsiveness of the Minimum Data Set," *J Amer Ger Soc,* Vol. 47, 1999, pp. 1000–1004.

Stineman M. G., J. E. Goin, B. B. Hamilton, and C. V. Granger, "A Case Mix Classification System for Medical Rehabilitation," *Med Care*, Vol. 34, No. 4, 1994a, pp. 366–379.

Stineman M. G., B. B. Hamilton, C. V. Granger, J. E. Goin, J. J. Escarce, and S. V. Williams, "Four Methods for Characterizing Disability in the Formation of Function Related Groups," *Arch Phys Med Rehabil*, Vol. 75, 1994b, pp. 1277–1283.

Stineman M. G., J. A. Shea, A. Jette, C. J. Tassoni, K. L. Ottenbacher, R. Fiedler, and C. V. Granger, "The Functional Independence Measure: Tests of Scaling Assumptions, Structure, and Reliability Across 20 Diverse Impairment Categories," *Arch Phys Med Rehabil*, Vol. 77, 1996, pp. 1101–1108.

Stineman M. G., A. Jette, R. Fiedler, and C. V. Granger, "Impairment-Specific Dimensions Within the Functional Independence Measure," *Arch Phys Med Rehabil*, Vol. 78, 1997a, pp. 636–643.

Stineman M. G., C. J. Tassoni, J. J. Escarce, J. E. Goin, C. V. Granger, R. C. Fiedler, and S. V. Williams, "Development of Function-Related Groups Version 2.0: A Classification System for Medical Rehabilitation," *Health Serv Res*, Vol. 32, No. 4, October 1997b, pp. 529–548.

Williams B. C., Y. Li, B. E. Fries, and R. L. Warren, "Predicting Patient Scores Between the Functional Independence Measure and the Minimum Data Set: Development and Performance of a FIM-MDS 'Crosswalk,'"*Arch Phys Med Rehabil*, Vol. 78, 1997, pp. 48–54.