July 2005

Development of Quality Indicators for Inpatient Rehabilitation Facilities

Final Report

Prepared for

CDR Rita Shapiro, M.A., P.T.
Project Officer
Office of Clinical Standards and Quality
Centers for Medicare & Medicaid Services
7500 Security Boulevard
Baltimore, MD 21244-1850

Prepared by

RTI International
411 Waverley Oaks Road - Suite 330
Waltham, MA 02452-8414

RTI Project Number 07964.004

Development of Quality Indicators for Inpatient Rehabilitation Facilities

Authors: Barbara Gage, Ph.D.
Shulamit Bernard, Ph.D.
Roberta Constantine, Ph.D.
Elisabeth Root, M.S.
Claudia Squire, M.A.
Jeremy Green, B.A.
Deborah Osber, M.P.H.
Brandy Gainor, B.A.

Margaret Stineman, M.D.
University of Pennsylvania

Project Director: Barbara Gage, Ph.D.

Scientific Reviewer: Leslie Greenwald, Ph.D.

Federal Project Officer: CDR Rita Shapiro, M.A., P.T.

RTI International*

CMS Contract No. 500-00-0024, T.O. #4

July 2005

This project was funded by the Centers for Medicare & Medicaid Services under contract no. 500-00-0024, T.O. #4. The statements contained in this report are solely those of the authors and do not necessarily reflect the views or policies of the Centers for Medicare & Medicaid Services. RTI assumes responsibility for the accuracy and completeness of the information contained in this report.

*RTI International is a trade name of Research Triangle Institute.

ACKNOWLEDGMENTS

This work is the result of many individuals' contributions. RTI would like to thank our esteemed colleague, Dr. Stineman, for the many hours she joined us in meetings with our CMS project team and her invaluable contributions throughout this process. We would also like to thank Anne Deutsch, who in addition to being a TEP member, led the hospital training effort for the primary data collection phase, contributed to the training manuals and provided the hands-on training for completing the revised forms. Our Technical Expert Panel members, Christine Baron, Grace M. Carter, Anne Deutsch, Bruce M. Gans, Carl V. Granger, Kurt Hoppe, June Isaacson Kailes, Peter A. Lichtenberg, Richard T. Linn, Samuel J. Markello, Kenneth J. Ottenbacher, Elliot J. Roth, Margaret G. Stineman, Deborah L. Wilkerson, and Carolyn Zollar provided invaluable directions, contributions, and insights throughout this process. They made themselves available for meetings and always responded quickly when asked for additional information.

This work would not have been possible without the hospitals that agreed to make their staff and units available to participate in the primary data collection: University of Pennsylvania Medical Center (UPMC), Philadelphia, PA; Rehabilitation Institute of Chicago (RIC), Chicago, IL; HealthSouth New England Rehabilitation Hospital, Woburn, MA; Baylor Institute for Rehabilitation, Dallas, TX; Spalding Rehabilitation Hospital, Aurora and Denver, CO; Pitt County Memorial Hospital—Regional Rehabilitation Center, Greenville, NC; St. Luke's Rehabilitation Institute, Spokane, WA; Cedars-Sinai Medical Center, Los Angeles, CA ; and Sunnyview Rehabilitation Hospital, Schenectady, NY. We thank them for taking the time and resources to contribute to this important policy effort.

We would like to thank the project officers, CDR Rita Shapiro who under the direction of Mary Pratt brought together a team of CMS participants with a wide range of valuable experience, including Jean Scott, James Bowman, Susanne Seagrave, Georgia Johnson, Pedro Diaz, and August Nemec. Their on-going participation in phone calls gave this project the continued attention it required.

We would also like to thank our programmer, Bob Baker, editor, Paul Angiolillo, and Documentation Preparation specialists, Nanci Pepoli, Norma DiVito, Linda Thompson, and Camille Quinn for their assistance in preparing this document.

CONTENTS

List of Tables

EXECUTIVE SUMMARY

The overall goal of this project, "Development of Quality Indicators for Inpatient Rehabilitation Facilities (IRF)," was to assist the Centers for Medicare and Medicaid Services (CMS) in developing appropriate measures to monitor and evaluate the quality of rehabilitation services provided to Medicare beneficiaries in IRFs. The key questions addressed by this report include: What are the expected outcomes of an inpatient rehabilitation stay? What factors affect those outcomes? Do we have appropriate measures of those relationships? And if not, what measures do we need? In considering these key questions, RTI incorporated expertise from the field of physical medicine and rehabilitation services to develop measures specific to the rehabilitation field that would target these issues while still minimizing the administrative reporting burdens for providers.

Purpose and Background

Inpatient rehabilitation facilities play an important role in the post acute continuum of care. These hospitals provide inpatient services to patients requiring acute level rehabilitation services and multidisciplinary rehabilitation treatment teams. Patients treated in these settings are receiving care for functional or cognitive impairments that have a likelihood of improving with treatment. Cases typically treated in these hospitals include those having experienced a stroke, spinal cord or brain injury, amputation, major multiple traumas, hip fractures, neurological disorders, burns or certain types of arthritis. Patients using this level of care need on-going monitoring or are significantly dependent in their activities of daily living but their functional levels are expected to improve with treatment. These patients have documented levels of "medical necessity" for inpatient levels of care, such as the need for 24 hour nursing and availability of physician services. They receive the equivalent of 3 hours therapy per day while in the IRF, a more intensive level of care than provided in a skilled nursing facility or general acute hospital.

In January 2002, the Medicare payment system for IRF services was modified to a case-mix adjusted prospective payment system (PPS). PPS payments are based on patient-level data collected in the inpatient rehabilitation facility patient assessment instrument (IRF-PAI). This new data collection activity, using the IRF-PAI for payment purposes, presented an opportunity for CMS to also collect information to monitor quality of care in the inpatient rehabilitation setting. To take advantage of this opportunity, CMS included interim quality items on the IRF-PAI. However, these items lacked consensus from the field, and therefore, were included as voluntary items until further work could be done to study their applicability with IRF populations. This project is intended to address that need.

The IRF-PAI tool grew out of years of work in the field of rehabilitation and disability research. It incorporates the Functional Independence Measures (FIM)^{TM} developed by Hamilton, Granger et al., during the 1980s, which were developed to measure patients' functional levels and evaluate patient improvements between admission and discharge from an IRF. In addition, the IRF-PAI collects mandatory information on patients' medical conditions, comorbid conditions, functional levels, insurance, and certain social support factors. While the current IRF-PAI data provides key pieces of information for understanding differences in patient severity levels and treatment needs, there are factors affecting outcomes that are not currently captured by the IRF-PAI. These include measures that are associated with different medical

conditions that need to be accounted for before evaluating the effects of facility care on patient outcomes. Due to the very limited time that CMS had to implement the first IRF-PAI instruments (time that was driven by mandatory implementation of PPS), the interim quality indicators included on the instrument were not pre-tested on the inpatient rehabilitation population, and instead were culled from other areas of healthcare, including nursing facility services. The purpose of this project was to examine these interim quality items on the IRF-PAI using an expert panel, a literature review, and empirical analyses of pilot test data.

Project Overview

Four main tasks comprise the majority of work completed by RTI under this project. The first task was a Technical Expert Panel (TEP) that provided input from a wide range of experts and industry stakeholders, well versed in the history, goals, and prior quality measurement efforts associated with treating inpatient rehabilitation patients and developing the FIM™ Functional Related Groups (FIM-FRG) system. The second task was an extensive literature review conducted by RTI on potential quality of care measures which might be modified or applied to an IRF patient population. The third task was an analysis of the quality of care data submitted voluntarily by some IRFs during the first year of the Federal data collection effort. Last, RTI developed and pilot tested a revised IRF-PAI instrument that included new items designed to form the basis of a quality monitoring system for Medicare.

The TEP, which was comprised of experts in physical medicine and rehabilitation services, including physiatrists, psychologists, physical, occupational, and speech and language therapists, nurses, rehabilitation researchers, quality improvement organizations, hospitals, associations, universities, consumers, and members of the UDS^{MR} team suggested two key outcome measures for monitoring inpatient rehabilitation services-- 1) changes in functional status and 2) discharges to community—the 2 key goals of an IRF stay. Using these outcomes as criteria, RTI re-examined the IRF-PAI data for appropriateness in measuring these concepts or risk adjusting the measurement of these two sets of outcomes. However, because of their interim, and therefore, voluntary reporting requirements, few IRFs elected to submit these data which made it difficult to assess their usefulness in the existing data set.

It was suggested that certain factors were critical in evaluating these outcomes and should be tested as possible additions to the IRF-PAI tool. These included measuring a patient's social support networks and their premorbid levels of physical functioning, cognitive functioning, depression, mood, and engagement in the therapeutic process. These factors directly impact the patient's ability to improve their functional status during inpatient rehabilitation and their likelihood of being discharged to the community. Having information on these patient characteristics could be important for monitoring the quality of care and understanding the potential impact of rehabilitation services for an individual.

Based on recommendations from the TEP, RTI conducted an extensive literature review on these quality of care concepts and specific items validated in other studies. The TEP evaluated the usefulness of these different instruments and identified those best suited for use in the IRF setting. These items were then added to a pilot instrument that was tested in a set of nine hospitals in late 2004 (October to January). While the results of this study are not nationally representative, the selected hospitals varied geographically, and in size, hospital affiliations, and

ownership. As a result, the instrument was tested in a wide range of organizations and across many IRF patient populations. By including a wide breadth of facilities and Medicare patient populations, RTI hoped to identify potential problems in using any of the items with particular groups.

The analysis of the pilot data suggest that many factors affect quality of care outcomes in the IRF setting. Different measures affect motor functional change than affect cognitive functional change or the probability of being discharged home. However, more work is needed to test these relationships on a national sample of IRF Medicare patients and to understand how these may vary for patients having different medical conditions. While the pilot sample is somewhat limited for effectively testing differences within impairment groups, this initial work suggests that differences exist and should be further studied. This is consistent with work by other members of the team, particularly Dr. Stineman, Dr. Granger, and their colleagues.

Results and Recommendations

The goal of this work is to provide CMS with a parsimonious set of indicators that achieve the goal of monitoring quality while minimizing the resources required to collect data. To that end, no items were included that would require every hospital to add a specific discipline (such as a psychiatrist or speech pathologist) to the treatment team for the sole purpose of collecting an item. In addition, the operational aspects of each item were taken into account. Participating facilities were asked to comment on both the usefulness and usability of each item in the pilot test. Having an item that met the conceptual goals but could not be scored reliably, for instance, would not contribute to a reliable set of data.

Modifications to the IRF-PAI instrument

The IRF-PAI instrument that was fielded in the pilot test included additions in a number of domains: pre-morbid social network, pre-morbid function, medical needs, and quality indicators. We tested multiple versions of certain items (swallowing, depression, respiratory status) to determine which worked best, if at all, in measuring IRF populations. In addition to selecting the best measure of those concepts, only those items that have contributed to the explanatory power of one of the outcomes models are included in the recommendations (see ***Appendix A*** for recommended form changes).

Several items were also removed from the original IRF-PAI. These include:

- Comatose and dehydration: less relevant for IRF populations;
- Clearing airways: poor measure for IRF populations;
- Balance: is difficult to define and measure in a standard manner; and
- Falls: expected during a rehabilitation process. Serious injuries, which should not occur, can be identified through the ICD-9 codes, making this item unnecessary.

As shown in ***Table E-1***, each of the variables discussed in the following recommendations were significantly associated with at least one outcome or were considered conceptually relevant; few were associated with every dependent variable.[1] This summary table shows the variability in the types of items that were associated with the total FIM motor change score, total FIM cognitive change score, and each of the decomposed FIM motor activities as well as the probability of discharge to the community. ***Table E-2*** shows the degree to which the addition of the new items increased the explanatory power of each of these models.

Based on these results, the domains, summary of key field test findings, and recommendations are as follows:

- Retain the pre-morbid social network questions involving residence and caregivers—this concept was well received by both the Technical Expert Panel and the field test sites. The belief that prior social network influences discharge to a community setting is one that is recognized as relevant.

- Retain the two questions currently on the IRF-PAI that collect information on pre-hospital vocations. While these questions are not relevant to the majority of the general Medicare population, they are relevant to the high number of younger disabled beneficiaries and their expected outcomes.

- Replace delirious with the assessment of orientation to person, place, and time. This item is understandable to the IRF staff, easy to gather, and relevant to IRF outcomes.

- Retain the current item for swallowing and consider additional research prior to including ASHA's Swallowing Functional Communication Measure in the IRF-PAI.

- Include on the IRF-PAI form a pre-morbid FIM score for each item; however, collapse to three rather than seven categories.

- Keep the FIM goals item. While it should not be used to measure quality due to its intended use as a planning tool rather than an assessment measure, some hospitals find it useful internally.

- Keep two of the three pain items; keep 50A which is part of existing IRF-PAI and add 50C which assesses the impact of pain on the patient's ability to participate in the therapeutic process.

[1] This summary is based on the results presented in Section 5, Tables 32 and 34.

Table E-1
Summary table of the multivariate models presented in Section 5

Independent Variables	FIM Motor Change	FIM Cognitive Change	FIM Self Care	FIM Sphincter Control Change	FIM Mobility Change	Discharge to Community. 1	Discharge to Community 2
Demographic Variables							
Gender (Female)				+		+	+
Married (Yes)							
Pre-Hospital Living Setting (Community)			+				
Age							
Severity of Illness							
Charlson Index					-		
Impairment Groups							
Central Nervous System		-					-
Spinal Cord Dysfunction	-	-	-	-		-	-
Neurologic	-		-			-	-
Musculoskeletal	-				-		
Stroke	-	-	-				
Endurance	-	-	-	-		-	-
Other						-	
FIM Items							
FIM Self-Care Adm	-		-	+	+		
FIM Sphrincter Control Adm		-		+		+	
FIM Mobility Adm			+	-	+	+	
FIM Cognitive Adm	-		+	+	-		
Premorbid Social Network							
Lives With							
Lives Alone							
Lives with Others							-
Assisting Persons							
PT Assist in Home		+					
PT Assist All Others						+	+
Assist Type							
Assist Type One							
Assist Type More than One							
Caregiver Expectations on Admission							
P CG Family Adm	-			-	-		
P CG Other Adm	-		-	-	-	-	-
Frequency of Caregiver on Admission							
Once daily or less	+		+				
Medical Needs							
Disoriented (Yes)			-				
Swallowing Problems (Yes)		-	-				
Shortness of Breath at Rest (No)							

(continued)

Table E-1 (continued)
Summary table of the multivariate models presented in Section 5

Independent Variables	FIM Motor Change	FIM Cognitive Change	FIM Self Care	FIM Sphincter Control Change	FIM Mobility Change	Discharge to Community. 1	Discharge to Community 2
Function Scores							
Premorbid Self Care Function				+			
Premorbid Sphincter Control Function		-				-	
Premorbid Cognitive Function	+	+			+		
Premorbid Mobility Function	+	+	+		+		
Quality Indicators							
Pain Limiting Rehab Adm.	+		+	+	+		
Most Sever Pain Rating Adm.						-	
Pain Treatment Modality							
One Pain Modality Adm.		+					
Two or More Pain Modality Adm.							
Pressure Ulcer							
One or more Pressure Ulcer Adm.							
Geriatric Depression Scale							
Patient Life Empty Adm. (Yes)						-	-
RIC FAS Scale							
Mood Problem Min. Adm.		+					
Mood Problem Mod. Adm.							
Mood Problem Severe Adm.		-					
Engagement							
Copes Problem Min Adm.		-				-	-
Copes Problem Mod. Adm.		-				-	-
Copes Problem Severe Adm.						-	-

NOTES: Only those relationships that are statistically significant are included in the table.
+ = Positive statistically significant effect on dependant variable
- = Negative statistically significant effect on dependant variable
Impairment Groups were relative to the Replacement of Lower Extremity

SOURCE: Section 5, Tables 32 and 34

Table E-2
Explanatory power of models without and with the additional Pilot test items (Adjusted R-squares)

	Existing Items		New Items
	National	Pilot	Pilot
FIM Motor Change	0.08	0.07	0.2
FIM Cognitive Change	0.24	0.32	0.52
FIM Self Care Change	0.12	0.16	0.3
FIM Sphincter Control Change	0.33	0.31	0.4
FIM Mobility Change	0.15	0.17	0.27

SOURCE: Section 5, Tables 30 and 32

- Keep the PUSH tool for pressure ulcers, but re-assess this item after collecting national data. Most hospitals document ulcers in some way. Having a measure of ulcer severity is important for identifying worsening ulcers. The other two ulcer questions should also be retained as they are important risk adjusters. Patients with ulcers will have lower participation in therapy.

- Replace the RIC-FAS depression item and the 4-item GDS with the Yale Depression Screen which is a one question item. The mood disorder screen may be a useful risk adjuster.

- Keep the RIC-FAS engagement item. Every hospital felt this was a key factor affecting outcomes and agreed that it should be considered in measuring quality of care.

- Re-align the discharge disposition question to map to the pre-morbid social network housing question.

Empirical Models and Analysis

As described in Section 5, the analyses performed on the Pilot Sample are exploratory**.** Interpretation of the results must take into consideration the small sample size, limited numbers of facilities participating in the pilot test, and the many questions and issues the hospitals raised regarding some of the piloted items being tested. Nevertheless, the outcomes of the analyses provide important direction for future work. Based on the results of the preliminary analysis, RTI's recommendations are:

- Condition-specific analyses need to be conducted using a larger sample—the limited field test precluded our ability to build risk-adjusted models that are condition specific. However, all our preliminary analyses demonstrated that the specific medical conditions leading to the rehabilitation stay significantly effected in outcomes. These additional analyses need to be performed using national data, given the small numbers of admissions for some of the impairment groups.

- Repeat analysis on a national level using a theory-driven technique, such as hierarchical modeling, where variables measuring domains are entered one at a time and their impact on the model is examined. The limited field test provided insufficient cases to examine the items within conditions. The utility and importance of these items will differ by condition and should be modeled separately. The analysis of field test data, by necessity, aggregated across hospitals and across conditions so that models need to be replicated using a larger pool of patients. Additionally, many of the pilot-tested variables were either aggregated into dichotomous variables or collapsed into fewer categories in order to test their effect in the models due to lack of sample size. Thus, larger samples will provide an opportunity to analyze the influence of multi-category variables more fully.

- We considered two major outcomes: discharge disposition and change in various FIM scores. Both cognitive and motor FIM scores were considered. Another method that can be used to conceptualize the outcome is to consider observed versus expected, condition-specific, risk-adjusted change scores (Stineman et al., 199X). Using this method would allow a hospital to track its performance relative to its case mix. We recommend that CMS undertake this type of analysis using national data.

Next Steps

This study was described at several national rehabilitation meetings this past year, either as part of a larger presentation on CMS' quality initiatives (Scott, 2004) or to professional rehabilitation research or provider audiences (Constantine, 2004; Gage, 2004; Deutsch, 2004). Interest was expressed in having national benchmarks that could be used to study individual hospital quality. Currently, there are several initiatives trying to use UDS^{MR} data, but not all hospitals participate in that data collection and the efforts are proprietary. Hospitals use these reports and develop other internal efforts to monitor quality for internal purposes. Where relevant, we built on these and the CARF-accreditation standards.

Much work remains to be done over the next few years to test these measures on the entire range of Medicare populations. Specifically, we recommend that CMS:

- Require hospital submission of the revised and recommended quality indicators. As noted in the current IRF-PAI analysis, few hospitals report voluntary items.

- Update IRFPAI training manual to include training materials on the new items. Also update the vignettes used in the training sessions to incorporate changes from the field test. Incorporate these changes into the existing training system, helpdesk, website and other resources.

- After at least one year of data:

  - Repeat and expand the data analyses using the techniques recommended above to examine importance of added items to risk adjusting outcomes.

  - Reassess the level of burden on facilities relative to importance of data collected.

  - Reconvene TEP to address results based on national dataset.

SECTION 1
INTRODUCTION

This study, "Development of Quality Indicators for Inpatient Rehabilitation Facilities (IRF)," is intended to assist CMS in developing appropriate measures to monitor and evaluate the quality of rehabilitation services provided to Medicare beneficiaries in IRFs. This task is particularly important as the field of inpatient rehabilitation services has undergone many changes in the last few years that introduce not only the opportunity for improving care by collecting more refined information on patient needs and outcomes, but also potentially new incentives to minimize care (Dobrez, LoSasso and Heinemann, 2004).

1.1 Background on Inpatient Rehabilitation Facilities, (IRFs), the IRF Prospective Payment System and Related Data Collection

Inpatient rehabilitation facilities play an important role in the post acute continuum of care. These hospitals provide inpatient services to patients requiring acute level rehabilitation services and multidisciplinary rehabilitation treatment teams. Patients treated in these settings are receiving care for functional or cognitive impairments that have a likelihood of improving with treatment. Cases typically treated in these hospitals include those having experienced a stroke, spinal cord or brain injury, amputation, major multiple traumas, hip fractures, neurological disorders, burns or certain types of arthritis. Patients using this level of care need on-going monitoring or are significantly dependent in their activities of daily living but their functional levels are expected to improve with treatment. These patients have documented levels of "medical necessity" for inpatient levels of care, such as the need for 24 hour nursing and availability of physician services. They receive the equivalent of 3 hours therapy per day while in the IRF, a more intensive level of care than provided in a skilled nursing facility or general acute hospital.

In January 2002, the Medicare payment system for these hospitals was modified to a case-mix adjusted prospective payment system (PPS). The goals of case-mix adjusted PPS are to more accurately reflect patient costs while also providing incentives for efficient patient care. This new risk adjusted PPS also introduced additional data collection activities that are required for payment, but which can also be used for monitoring quality of care in these settings. The implementation of this new PPS system, and the data available through it, provide an opportunity for improving CMS methods to measure and monitor quality of care in the inpatient rehabilitation setting.

The key data initiative associated with the IRF PPS is the mandatory submission of patient assessment data. These data provide information on patients' medical conditions, comorbidities, functional levels, medical needs, insurance, and certain social support factors. Hospitals are now required to submit data using the inpatient rehabilitation facility patient assessment instrument (IRP-PAI) which incorporates the Functional Independence Measures (FIM)^{TM} developed by Hamilton, Granger et al., during the early 1980s. The IRF-PAI tool grew out of years of work in the field of rehabilitation and disability research developing and applying the FIM™ measures of patients' functional ability levels to evaluate patient improvements between admission and discharge from an IRF. Many of the patient assessment items included in

the IRF-PAI were already being collected by a large number of rehabilitation hospitals as part of their participation in the Uniform Data System for Medical Rehabilitation (UDS^{MR}).

The Uniform Data System for Medical Rehabilitation (UDS^{MR}) was developed in the mid-1980s by a task force consisting of representatives from the rehabilitation community (Hamilton, Granger, et al., 1987). Under the leadership of Carl V. Granger and Byron B. Hamilton, and a grant from the National Institute for Disability and Rehabilitation Research, U.S. Department of Education, the task force was charged with developing a uniform way of documenting the severity of disability, including both cognitive and motor-related disabilities, and the outcomes for medical rehabilitation that was valid, reliable, and sensitive to change.

Beginning in 1988, a large number of rehabilitation facilities submitted patient data using the Uniform Data System to a national repository based at the University at Buffalo, The State University of New York. Since that time, the data set and training materials have been updated (Version 5.1 was in use until 2002), and studies of the reliability (Hamilton, Laughlin, et al., 1994), validity (Stineman, Shea, et al., 1996; Linacre, Heinemann, et al., 1994), and responsiveness (Dodds, Martin, et al., 1993) of the FIM™ instrument have been conducted. In addition these data were used to develop the Functional Independence Measure- Function Related Groups (FIM™-FRGs), subsequently renamed Case Mix Groups (CMGs). The FIM™-FRGs system was developed by Stineman et al. during the early 1990s to measure and quantify ADL functions and functional status with respect to basic personal care (Stineman MG, Escarce JJ, et al., 1996; Stineman MG, Tassoni CJ, et al., 1997). The FIM™ instrument represents the hours of care by another person that may be needed if functional improvement does not occur. It was developed to quantify severity of disability. Externally tested and revised by Carter et al. (1997), the FIM™-FRGs form the patient classification system used in the IRF PPS (Carter, Relles, et al., 1997).

Prior to the Medicare IRF PPS, many IRFs collected UDS^{MR} data on all inpatients (regardless of payer) and voluntarily subscribed to national databases in order to receive feedback reports that describe the characteristics and clinical outcomes of patients discharged from their IRF as well as contrasting regional and national statistics based on the submitting facilities. Administrators at participating facilities use the information from these reports for internal quality management efforts, in newsletters and marketing materials, and to help meet accreditation requirements.

When the Uniform Data System for Medical Rehabilitation, including the FIM™ instrument (Version 5.1), was incorporated into the Medicare IRF-PAI, some modifications were made, including:

- The assessment time frame was changed to cover the entire first 3 days of a patient's stay for the admission assessment, and the entire period covering the last 3 days for the discharge assessment. The discharge assessment was further updated in April 2004 to cover any 24-hour period during the last 3 days.

- The definitions for two FIM™ items (Bladder Management and Bowel Management) were updated to include a look-back period of 7 days. (This look-back period had previously covered up to 2 weeks.)

- The "0" category on the 7 point FIM™ scale was added to note that an activity does not occur at admission.

- Functional modifiers were added to the data set in order to assist with more accurate data collection of certain FIM™ items.

- Medical Needs and Quality Indicator items were added but not tested on the IRF populations.

The initial IRF-PAI forms collected information on beneficiary functional status, data that are not available on claims but are considered an important measure of rehabilitation treatments. These types of data provide key pieces of information for understanding differences in the severity levels, treatment needs, and expected outcomes of individuals having different medical conditions or being treated in different facilities. In addition to data on the patient's medical condition, insurance information, and pre-admission residential and vocational items, an interim set of medical need and quality indicators were included. Because of the very limited time that CMS had available to implement the first IRF-PAI instruments (time that was driven by mandatory implementation of PPS) these interim quality indicators were not pre-tested on the inpatient rehabilitation population and instead were culled from other areas of healthcare, including nursing facility services. Because of their interim nature and because they are not required to implement the IRF PPS, these quality items added by CMS to the IRF-PAI instrument are not part of the mandatory data set that must be submitted by all IRFs. In practice, many IRFs have elected not to submit these voluntary quality-related data elements.

1.2 Overview of the RTI Study

This report reviews the voluntary IRF-PAI items, the frequency with which they have been submitted by IRFs, and their appropriateness for use as quality monitoring items. We also address whether other areas should be considered for monitoring facility-level quality of care. In addition, in response to a specific request by CMS, we consider whether the available data would allow CMS to risk adjust, or control for patient differences, across facility populations when evaluating outcomes.

This report reflects findings from four main tasks that comprise the majority of work completed by RTI under this project. The *first* task was a Technical Expert Panel (TEP) through which we were able to consult with a wide range of experts and industry stakeholders, well versed in the history, goals, and prior quality measure development efforts associated with the treatment of patients receiving inpatient rehab and development of the FIM™ Functional Related Groups (FIM™-FRG) system. Building on the past efforts to develop FIM™-FRG-based quality measurement for inpatient rehabilitation services (Stineman, 1998; Stineman, 1995), RTI added analyses of the first two years of national IRF-PAI data. This was done to assess the usefulness of the voluntary items and to generate consideration of the appropriateness of the IRF-PAI and Medicare claims for monitoring quality. The *second* task was an extensive literature review conducted by RTI on potential quality of care measures used in non-IRF settings, but which might be modified or applied to an IRF patient population. The *third* task was an analysis of the voluntary interim quality of care indicators submitted by some IRFs through the current IRF-PAI

instrument. The *fourth* task was the development and pilot testing of a revised IRF-PAI instrument that included new data collection items that could form the basis of a quality of care monitoring system for IRFs.

Key outcomes for inpatient rehabilitation services used in this report were identified initially by our Technical Expert Panel (TEP). The TEP for this project was comprised of experts in physical medicine and rehabilitation services and included physiatrists, psychologists, physical, occupational, and speech and language therapists, nursing, rehabilitation researchers, quality improvement organizations, hospitals, associations, universities, consumers, and members of the UDS^{MR} team. The TEP members suggested that key outcomes expected from an inpatient rehabilitation stay should include *changes in functional limitations* and *discharges to community*—the 2 key goals of an IRF stay. Using these outcomes as criteria, the TEP re-examined the IRF-PAI data for appropriateness in measuring these concepts or risk adjusting the measurement of these two sets of outcomes. It was suggested that certain factors were critical in evaluating these outcomes and should be pilot tested as possible additions to the IRF-PAI tool. These included measuring a patient's social support networks and their premorbid levels of physical functioning, cognitive functioning, depression, mood, and engagement in therapy. These factors directly impact the patient's ability to improve their functional status during inpatient rehabilitation and their likelihood of being discharged to the community. Having information on these patient characteristics could be important for monitoring the quality of care and understanding the potential impact of rehabilitation services for an individual.

Based on recommendations from the TEP, RTI conducted an extensive literature review on quality of care concepts and specific items in use in other settings to identify data collection items validated in other studies. The TEP evaluated the usefulness of these different groups of instruments and identified those best suited for use in the IRF setting. These items were then added to a pilot instrument that was tested in a set of nine hospitals in late 2004 (October to January). While the results of this study are not nationally representative, the hospitals were selected to vary geographically, by size, ownership and whether they were hospital units or freestanding facilities. This provided a wide range of populations and organizational characteristics in the facilities that tested the pilot instruments. By including a wide breadth of facilities and Medicare patient populations, we could identify potential problems in using any of the items with particular groups.

The underlying literature review, TEP discussion, and results of the pilot tests are presented in this report. The analyses suggest that the factors that predict quality of care outcomes in the IRF setting vary. Different measures affect the extent of motor functional change than affect cognitive functional change or the probability of being discharged home. Therefore, the results presented in this report are important to begin understanding the factors that may affect IRF Medicare patient outcomes and the quality of their care. However, more work is needed to test these relationships on a national sample of IRF Medicare patients and to understand how these may vary for different types of patient conditions. While the pilot sample is somewhat limited for effectively testing differences within impairment groups, this initial work suggests that differences exist and should be studied further. This is consistent with work by other members of the team, particularly Dr. Stineman and her colleagues.

Development of quality measurement items for inpatient rehabilitation services can contribute to a number of CMS responsibilities, including near-term evaluation of the effects of IRF PPS on quality of care. CMS has a long history of this type of program monitoring and evaluation; the agency has monitored quality and access to care after the implementation of each PPS system from hospital DRGs in 1983 to the more recent implementation of the home health and SNF PPSs in the late 1990s. Many of the RTI team members, including the wide panel of consultants and TEP members, have been involved in these past efforts. The resulting work will allow CMS to develop IRF-specific performance measures supported by the field and recognizing the types of contributions inpatient rehabilitation facilities provide to improving Medicare beneficiaries' access to appropriate health care. This work also contributes to CMS long-term goals of identifying and providing incentives for provider quality improvement. Without these IRF-specific measures, CMS would be limited in its ability to measure, track, and set standards for improvement in inpatient rehabilitation services. The results of this study will be useful to CMS, the hospitals, and consumers choosing among different treatment options. RTI's recommendations are presented in the final section to identify future directions for continuing this important endeavor.

1.3 Organization of the Report

This report is organized into 6 sections. These sections reflect the process through which the work developed. The first section is the introduction and provides an overview of the project goals. The second reports on the Technical Expert Panel meetings, the data presented at those meetings, and their recommendations for developing outcome measures. This section also includes a recommendation to modify the IRF-PAI and conduct a pilot test using the revised tool. The third section presents the literature review that the TEP members used to select additional items for the IRF-PAI. The fourth section summarizes the primary data collection effort and reports descriptive statistics on the proposed additional IRF-PAI items. The fifth section discusses the use of the primary data set, in conjunction with validated mandatory items from the IRF-PAI that were submitted for the pilot sample, to analyze the usefulness of these new measures in predicting outcomes. The sixth section contains RTI's recommendations, including suggestions made by the TEP members after reviewing the results of this study.

SECTION 2
TECHNICAL EXPERT PANEL MEETINGS

2.1 Purpose of the Technical Expert Panel (TEP)

The primary task of this contract was to identify quality indicators pertinent to the inpatient rehabilitation setting and to determine what information is necessary to calculate those indicators. To this end, the contract called for the formation of a panel of nationally recognized experts in the field of rehabilitation medicine. The panel included major stakeholders in the field, including hospital administrators and clinicians, clinical researchers, industry associations, and patient advocates. RTI also sought to involve individuals with the following clinical expertise: physiatry (physicians specializing in physical medicine and rehabilitation), nurses, physical therapy (PT), occupational therapy (OT), and speech and language therapy (ST). The meetings were particularly important for ensuring that perspectives specific to the inpatient rehabilitation community were incorporated into the work of this project. The results of both the first and second Technical Expert Panel (TEP) meetings are summarized here and are available in several other reports (Keller, Greenwald, and O'Keefe, 2002; Bernard, Root, and O'Keefe, 2004).

2.2 Method of Selecting the TEP

Potential Technical Expert Panel members were identified in conjunction with CMS by RTI staff with input by RTI's expert consultant to the project, Margaret Stineman. Dr. Stineman has an international reputation for her work in physical medicine and rehabilitation and outcomes. A large list of potential TEP members was generated, and they were contacted in November 2001, 3 months before the meeting was to be held. We required that the TEP members participate in person and that they be able to commit to both the first (2/2002) and second (2/2003) meetings. Some potential members were disqualified because they could not meet these criteria. The composition of the final TEP membership and their qualifications are summarized in biosketches presented in ***Appendix B***. This appendix also illustrates the range of the TEP membership, which includes industry and patient representatives, physical therapy, speech therapy, and cognitive functioning experts, hospital administrators, clinical researchers, psychiatrists, and nurses.

We identified more potential TEP members than the project could accommodate. In order to benefit from the contributions of these experts, we invited them to be "listeners" at the first TEP meeting. ***Appendix C*** includes the list of invited listeners.

2.3 First Technical Expert Panel Meeting

The first Technical Expert Panel meeting was held on February 20, 2002 in a conference room within the Humphrey Building in Washington, DC. The agenda for the TEP meeting is included in ***Appendix D.*** The meeting was transcribed by ACE-Federal Reporters, Inc. and an edited version of the transcripts was submitted to CMS (Transcript, 2002). During the 1-day meeting, personnel from RTI and CMS clarified the purpose of the contract and the TEP members' role, and led TEP members in discussing several key points:

1. the relevance of the current optional items on the IRF-PAI instrument for measuring quality of care,

2. additional domains (and measures) that should be included in the IRF-PAI,

3. other efforts currently underway to identify/develop measures for quality of care in physical medicine and rehabilitation, and

4. potential databases that could be used to construct and test quality measures.

The IRF-PAI has both mandatory and voluntary reporting items. Most of the mandatory items are used in the IRF PPS and include information on the beneficiary's insurance, pre-admission residence and vocation, medical conditions, functional impairments, and discharge-related information. The voluntary items included the medical needs section and the quality indicators. The TEP focused its attention on the quality indicators section.

Initially, TEP members were concerned about the appropriateness of certain IRF-PAI items that had been imported from the nursing home setting but were seen as less appropriate for IRF patients, such as those in the medical needs section. This included measures of being comatose, delirious, and dehydrated. Many of these factors were added to the IRF-PAI when the PPS was being established and were taken from the MDS form used in the nursing facility PPS. The ensuing discussion focused on the differences between rehabilitative medicine and nursing home services, and whether some of these items that had come from the nursing home community would affect patient outcomes or a patient's ability to participate in the rehabilitation process. In some cases, such as the delirium item, the TEP agreed that a concept was important to include but that a more appropriate measure, such as orientation, should be used for the inpatient rehabilitation services population. Other items originally from the MDS were suggested to be irrelevant to the rehabilitation populations or too difficult to objectively measure, and therefore, of limited use. For example, the TEP felt that few patients would be admitted to an IRF in a comatose condition because they would not be able to undergo three hours of therapy a day, engage in intensive therapy in a meaningful way, or be in the IRF long enough to demonstrate improvement given the short average length of stay in these hospitals.

In deliberating on each item, the panel also considered whether performance of the potential measure could reasonably be affected by the facility being held accountable: Would it be reasonable to judge facility performance based on the measure? In some cases, the panel felt that items were important to collect but should be used for risk adjustment rather than outcome measurement purposes. For example, it would be difficult to use pressure ulcers as an outcome measure because frequently patients are admitted from the acute hospital with the ulcer. Moreover, the etiology of an ulcer becoming apparent in rehabilitation might have been caused by an earlier event. While the ulcer should not worsen or have further breakdown in the IRF, it would be unfair to ascribe poor care by the IRF because of its presence. It is important, however, to account for the ulcer in assessing changes in functional status or discharge to a community setting because the presence of the ulcer may limit the patient's ability to fully engage in therapy.

The TEP also suggested several new domains not currently included in the IRF-PAI and suggested measures for each of these domains. This section summarizes the recommendations made during the first TEP meeting. The final report of the first TEP meeting was submitted to CMS in an earlier deliverable (Keller, Greenwald, and O'Keefe, 2002).

2.3.1 Evaluating Current IRF-PAI Items

The TEP reviewed each of the items in the voluntary sections of the IRF-PAI, including Medical Needs with its 4 items (comatose, delirium, dehydration, and swallowing), the three respiratory status indicators (difficulty clearing airways and shortness of breath with exertion and at rest), Pain ratings, 3 Pressure Ulcer items (staging, number, and the PUSH tool), and 2 Safety measures (balance and falls) to discuss how each could be used as either an outcome or a risk factor in monitoring facility-level quality of care.

Pain—In general, the TEP agreed that pain was one of the most important of these items in terms of affecting patient outcomes and that it should be measured at both admission and discharge, although for different purposes. At admission, a measure of pain severity should be used as a risk adjuster. Pain levels may affect a patient's ability to actively engage in therapy, and therefore reduce the functional improvements, which, in turn, may also delay the patient's ability to return home. The discharge measure is useful in looking at change in pain, which could be used as an outcome indicator. In addition, measuring the degree to which pain interferes with activity at both admission and discharge would be useful for risk adjustment purposes in estimating expected outcomes from the IRF treatments. The TEP suggested that RTI investigate the best measure of pain for these applications. In addition, consideration should be given about how to measure pain among patients with lower cognitive abilities who are unable to give an accurate pain report.

Swallowing—All TEP participants agreed that the swallowing function is important to rehabilitation. Most felt it should be a risk adjuster because those with problems swallowing are more difficult to treat – a danger of aspiration pneumonia is always present no matter what is done clinically. However, speech-language pathologists would also view swallowing as an outcome, that is, as a functional domain that could be improved through therapy.

Difficulty Clearing Airways—The TEP agreed that difficulty clearing airways was important to rehabilitation as a risk adjuster for reasons similar to the swallowing indicator. Patients with problems clearing airways have a higher probability of getting aspiration pneumonia regardless of clinical care. TEP members felt that an indicator of difficulty clearing airways would help to get at the severity of underlying pulmonary dysfunction (such as chronic obstructive pulmonary disease—COPD), which would affect the quality of therapy.

Shortness of Breath—The TEP agreed that dyspnea would impact progress in rehabilitation, but that it should not be included on the IRF-PAI because it could not be measured objectively by clinician reports. The only objective way to measure dyspnea is a pulmonary function test, and not all facilities have the equipment to complete this test. In addition, TEP members were concerned about the possibility of gaming if the indicator is used for risk adjustment.

Frailty—The TEP agreed that frailty was crucial to progress in rehabilitation. However, the measurement of frailty, as well as the definition, is complicated by a lack of agreement among experts. The field of geriatrics does not have consensus on its definition or its measurement. The TEP strongly recommended that RTI explore the definition and measurement of frailty.

Cognition-Related Domains—The TEP discussed a wide variety of cognitive states that could interfere with progress in an IRF. They emphasized that "mental fragility" was a particular concern for the Medicare population, but that sometimes it is treatable (when the cause is known) and sometimes it is not. The TEP made four distinct recommendations for cognitive-related domains:

1. The coma item should be deleted from the IRF-PAI. They felt that as a general indicator of quality, coma would be hard to define. It also was an unlikely condition for a patient admitted to an IRF.

2. The TEP agreed that delirium should be included as a risk adjuster since it complicated the facility's ability to provide treatment, but the delirium item on the current IRF-PAI should be replaced with a better measure.

3. The TEP recommended that any cognitive measures added to the IRF-PAI should provide information in addition to that available through the cognitive Functional Independence Measure (FIM™). The cognitive FIM™ taps the lower levels of cognitive functioning. A good supplement to the FIM™ would be measures of higher-level cognitive functioning. Some TEP members suggested the use of IADLs as a measure of higher-level functioning that could be used to supplement the FIM™. IADLs, such as shopping or using a telephone, tap into cognitive skills, but also depend on physical abilities. Consequently, they cannot be seen as a pure measure of cognition..

4. The TEP suggested adding items to measure depression and attention deficit. These were two aspects of cognitive functioning they felt could impact or interfere with the ability to benefit from therapy.

Pressure Ulcers—There was a consensus that the stage and number of pressure ulcers was important as both a risk adjuster and an outcome that should be measured at admission and discharge. The TEP agreed that the presence of pressure ulcers could influence participation in rehabilitation and that progression of pressure ulcers might indicate poor quality of care. However, the TEP members felt that the PUSH tool was not relevant for many of the cases treated in IRFs.

Patient Safety Indicators—The IRF-PAI contained two safety indicators – one measured balance problems at admission and discharge and the other counted the number of falls during an inpatient stay. The TEP felt that recording the number of patient falls was not a useful indicator of patient safety because rehabilitation patients may be expected to fall when they are working to regain function. The TEP agreed that the issue of hospital-acquired injury was more pertinent. Thus, a more legitimate measure would be the number of falls that resulted in serious harm, such as death, fracture, or laceration. These are measures that can be identified through the codes on the Medicare claims and are not needed here. Related to the issue of safety and falling were the IRF-PAI questions that addressed dizziness and balance. The TEP agreed that these states were

not important in and of themselves, but the issue was whether these states led to preventable injury. Hence, in the interest of keeping items only if they were either needed for payment or quality measurement, the panel agreed that dizziness and balance questions should be deleted from the IRF-PAI.

2.3.2 Domains That Should Be Added to the IRF-PAI

In addition to evaluating the value of existing IRF-PAI items for monitoring quality of care, the TEP members were asked to identify other factors that would be important quality indicators or that needed to be controlled for in estimating facility-level outcomes. Several areas were raised

Premorbid Functioning—The TEP agreed that premorbid condition (e.g., degree of physical functioning before the acute event that lead to the IRF admission) has an important effect on rehabilitation outcomes and must be used as a risk adjuster. No specific measures were suggested for this domain.

Depression—The TEP agreed that patients who suffered from depression would have poorer outcomes in the IRF. While the TEP agreed that the typical IRF length of stay (LOS) is not long enough to witness an improvement in endogenous depression, they felt it was important to measure for two reasons: it would affect a patient's ability to improve and therefore should be used as a risk adjuster and, secondly, it would encourage a good process of care. However, there was less certainty among TEP members regarding how depression should be measured because some clinicians would accept no less than a full psychiatric evaluation. They suggested using the word "mood" rather than "depression," as mood could be measured for these purposes using a single item such as the "downhearted and blue" item from the SF-36. Other members suggested using the Geriatric Depression Scale (GDS) or the CDC Epidemiological Survey of Depression (CES-D).

Patient Satisfaction—Many TEP members felt very strongly that patient satisfaction should be added to the quality outcomes monitored. Patient satisfaction is important to IRF administrators and was an important outcome for a wide variety of respondents in the Commission on the Accreditation of Rehabilitation Facilities (CARF) quality indicator study. However, the TEP decided against including satisfaction in the IRF-PAI tool as it was felt that this instrument was not an appropriate vehicle for collecting this type of patient-reported information. Other initiatives are underway in the Medicare program to collect patient satisfaction surveys in hospitals, and the TEP felt using these vehicles in the rehabilitation hospitals would be a better source of this information than adding an item to the IRF-PAI.

Other Quality Process Indicators—The TEP emphasized the importance of including healthcare process measures in the IRF outcomes monitoring system because (1) they were valuable in their own right as quality indicators and (2) process measures were strongly related to outcome measures. Process indicators could include providing patients with high-quality equipment while in rehabilitation. Such equipment would prevent injury (pressure ulcers) and enable patients to spend more time in therapy.

2.3.3 First TEP Meeting Follow-Up

From the initial TEP meeting, participants provided RTI with valuable suggestions on which IRF-PAI items should be deleted, items that were valid but required revision, and additional domains that should be added when the instrument is revised. As a result of the TEP suggestions, CMS requested that RTI conduct a broader survey of potential quality indicators, including those potentially beyond the scope of the current IRF-PAI instrument. As a result, an extensive literature review was conducted (see Section 3). The TEP suggested that RTI consider the definition and measurement of several additional domains listed below, and RTI used these recommendations to guide the literature review.

- Depression
- Premorbid functioning and social network
- Pain
- Physical Functioning
- Cognitive Functioning

2.4 Second Technical Expert Panel Meeting

The second Technical Expert Panel meeting was held November 6 and 7, 2003 in a conference room at the Centers for Medicare & Medicaid services in Baltimore, MD. The agenda for the TEP meeting is included in ***Appendix E***. The meeting was transcribed by Gore Brothers Reporting and Video Company, Inc. and an edited version of the transcripts are available at CMS (Transcript, 2003). The goal of the TEP meeting was to gather participants' feedback on several domains not measured by the current IRF-PAI. During the 2-day meeting, personnel from RTI and CMS led TEP members in discussion of several topic areas:

- Premorbid social supports and networks
- Physical functioning
- Cognitive functioning
- Depression or "mood" and engagement
- Pain
- Swallowing
- Process measures
- Outcome measures in the rehabilitation setting

Initially, TEP members expressed concern that the agenda did not include a discussion of specific outcome measures that are meaningful to inpatient rehabilitation. These included measures such as discharge back into the community and functional improvement. These items were subsequently added to the agenda. In addition, several TEP members felt that during the first TEP meeting, more emphasis was placed on covariants or risk-adjustment than on quality measures. At the second meeting, the TEP agreed to explicitly address quality measures, such as the recovery of physical and cognitive function. TEP members discussed the relevance of each topic area and, when appropriate, suggested items that could be used to measure quality or risk-adjustment. When panel members could not supply a specific item or instrument, they suggested organizations that may be able to provide those items.

2.4.1 Discussion of Additional Domains for the IRF-PAI

Premorbid Social Supports and Networks—The TEP agreed that some measure of social support should be included for risk adjustment in the IRF-PAI instrument, because it may give people with more challenging social situations a better chance at being admitted to rehabilitation. Members suggested that the items currently on the IRF-PAI measuring prehospital living setting/living with and prehospital vocational category/effort should be kept, but that an additional item was needed to assess environmental considerations — family and social resources available to the discharged patient. The additional item should address whether patients have the amount and quality of social supports they will need once they leave rehabilitation. In addition, members thought the question should be asked at discharge, to assess what resources will be available in the setting the patient is going into.

TEP members suggested that RTI follow up by looking for instruments that already have measures of social support. Two specific instruments suggested were the rehabilitation indicators used by the Commission on the Accreditation of Rehabilitation Facilities (CARF) and those in CMS' Outcome and Assessment Information Set (OASIS).

Physical Functioning—The TEP agreed that the level of physical functioning prior to the acute event is important for risk adjusting change in the Functional Independence Measure (FIM™) scores and should be included in the IRF-PAI instrument. Panel members considered several scales, but decided that any scale used to measure physical functioning should be comparable to scores obtained by the FIM™ so data can be compared at admission and discharge. The ability to compare data could lead to a new quality measure: How closely did a patient return to premorbid functioning level? TEP members agreed that several questions on the FIM™ could be asked in a premorbid context (i.e., before the current acute event that led to rehabilitation). They suggested adapting the FIM™ by constructing an additional question that assesses the patient's level of functioning (in each FIM™ domain) before the episode that led to hospitalization and the IRF admission. TEP members asked RTI to construct several items that could be used to adapt the FIM™ and contact panel members for input on which one(s) should be included in the revised instrument.

Cognitive Functioning—TEP members discussed possible cognitive tasks or items that could be abstracted from existing cognitive measures and added to the IRF-PAI which would predict outcomes independently of the cognitive items on the FIM™. Panel members suggested that the evidence shows that assessing executive cognitive function (ECF) in rehabilitation settings would add information that is not currently collected by the IRF-PAI. Members stressed that any

measure included on the IRF-PAI would not replace neuropsychological testing but instead improve the utility of the IRF-PAI. However, there was disagreement over what type of measure (e.g., direct measure of ECF or performance-based measures) would be most appropriate for the IRF-PAI. There was general agreement that this project should use cognitive impairment as a risk-adjuster only, not as an outcome measure at this time.

Since there was little agreement on how best to measure cognitive functioning, TEP members asked RTI to review some of the everyday functional cognition tasks that are part of other instruments and scales and to suggest 1 or 2 specific items to the TEP for consideration. If TEP members can reach agreement about an item, then it could be included in the instrument during the field test.

Depression—The TEP agreed that depression was important for risk adjustment because patients who are clinically depressed have poorer outcomes. However, the TEP cautioned against including improvement in clinical depression as an outcome, arguing that adequate assessment of depression through a screening item is not feasible. The TEP suggested it might be better to use a measure of "mood" rather than "depression" for risk adjustment. They cautioned that mood scales cannot diagnose depression, but they can be used as a screening tool to identify patients with possible mood disorders who need referrals to an appropriate professional for a full assessment. Several TEP members suggested creating a process measure to assess whether a patient who screened positive for a mood disorder received the appropriate referral and subsequent psychiatric assessment. The TEP asked RTI to identify items for assessing mood and solicit feedback before field testing an item.

TEP members also suggested adding to the IRF-PAI a general measure of engagement in the therapeutic process as a risk-adjuster. Members agreed that there are many different causes of non-participation—depression, anxiety, low arousal—all of which can interfere with the patient's ability to benefit from therapy and ultimately lead to poorer outcomes. Members agreed that it was not necessary to identify the specific cause of non-participation. Therefore, a broad measure of engagement would allow IRFs to account for a wide range of factors that influence a patient's level of engagement. The TEP asked RTI to investigate measures of engagement that have already been developed and suggested CMS' Minimum Data Set (MDS) activity items and Rehabilitation Institute of Chicago Functional Assessment Scale (RIC-FAS) items for ability to participate in therapy.

Pain—The TEP agreed that the pain item currently on the IRF-PAI should remain as a mandatory item and be used for risk adjustment. Most hospitals use this item, and moving to a different item would only create additional burden for them. Members agreed that pain is not a good outcome measure for rehabilitation because therapy often causes a certain amount of pain. It is an important risk adjuster, however, because having pain and taking medications for it can interfere with other functions.

Swallowing—TEP members agreed that swallowing should be a mandatory item on the IRF-PAI. Patients with swallowing disorders require a much higher level of care, making this a potentially important risk adjustment item. Panel members agreed that using swallowing ability for risk adjustment was an empirical question and one that was important to explore. They recommended testing the swallowing item currently on the IRF-PAI as well as an additional item from the American Speech-Language and Hearing Association's (ASHA) National Outcomes

Measurement System (NOMS). TEP members did not feel swallowing should be used as an outcome measure as it may create an incentive to remove a feeding tube too early, which may encourage inappropriate treatment decisions.

Process Measures—TEP members agreed that process measures should be included in the IRF-PAI as measures of quality. Process measures are designed to measure administrative, clinical, and utilization processes that are believed to lead to positive quality outcomes. The use of process measures is based on the belief that improving the process of care will yield improved outcomes. Evidence-based process measures are used in areas such as acute care because the evidence already exists pointing to specific process that leads to the desired outcome. There are far fewer studies linking rehabilitation processes to outcomes. Consequently, the effects of particular processes are less clear in the rehabilitation setting. TEP members agreed that to engage in a discussion of process measures and generate a list of potential measures would be outside the scope of this project. However, all felt some measures of process should be included, and suggested that RTI look at existing practice guidelines and expert panels that have already constructed process measures in a number of different areas. Because process measures differ depending on diagnosis, TEP members provided RTI with a list of the most prevalent conditions in rehabilitation that should receive specific consideration:

1. Stroke
2. Lower extremity joint replacement
3. Hip fracture
4. Brain injury
5. Spinal cord injury
6. Amputation

2.4.2 Outcomes in the IRF Setting

At several points during the session, TEP members raised the concern that we needed to identify outcomes that are important in the rehabilitation setting. Several argued that measuring quality is difficult when the desired outcomes are unclear because creating measures of quality requires researchers to work backward. The TEP generated a list of six outcomes thought to be appropriate to inpatient rehabilitation:

1. Discharge to the community
2. Level of functional independence
3. Patient satisfaction and experience with care
4. Patient satisfaction with quality of life
5. Change in functioning relative to admission
6. Medical stability

2.4.3 Preliminary Analysis of IRF-PAI Data

To better understand the current and potential use of the voluntary items on the IRF-PAI, RTI presented a preliminary analysis of the first 1.5 years of IRF-PAI data to the TEP.[2] The presentation was intended to 1) provide information regarding the completeness and quality of the Inpatient Rehabilitation Facility Patient Assessment Form (IRF-PAI) data in order to assist the TEP in examining the potential use of this data for quality indicators and 2) obtain input on direction for future analyses.

The data represented Medicare admissions beginning with the collection of the IRF-PAI data (January 2002) and discharge dates no later than July 31, 2003. The total number of admissions included in the analysis was 696,201, or 93% of total Medicare admissions. The following categories of cases were excluded in the analyses:

- test cases
- railroad beneficiaries
- admissions prior to January 2002
- implausible discharge dates
- duplicate stays
- unmatchable Beneficiary Identification Codes

The Medicare program was the primary payer for 98% of the admissions analyzed, and the findings were reported separately for Medicare primary and secondary cases during the TEP presentation. For this report, we focus on data for admissions when the Medicare program was the primary payer. The preliminary analysis examined all major sections of the IRF-PAI form for completeness and provided descriptive data for selected items. The presentation to the TEP included the following topics: 1) overview of each section of the IRF-PAI form, 2) description of mandatory and/or voluntary items within each section, 3) completeness of data collection and descriptive data for selected variables, and 4) overview of next steps for the analysis. This section summarizes the findings of the analysis.

Patient Information—The data demonstrated that the average age of Medicare patients admitted to inpatient rehabilitation facilities was 74 years of age, median 76, with a standard deviation of 10 years. Females made up the majority of inpatient rehabilitation admissions, representing 63% of total admissions. The ethnicity of the rehabilitation admissions varied as follows: 84.3% white, 9.5% black or African American, 3.4% Hispanic or Latino, and 1.5% other (Asian, American Indian, Alaskan Native, other Pacific Islander). Of Medicare patients admitted to inpatient rehabilitation facilities, 43.7% were married, 39% widowed, 7.7% never married, 7.3% divorced, and 1.4% with marital status unknown or not reported.

---

2 RTI did not have access to the IRF-PAI data until September 2003.

Initial rehabilitation stays made up 94.2% of all Medicare admissions, 4.2% of stays were classified as readmissions, and 1.6% were unplanned discharges, evaluations, or continuing rehabilitation stays. The vast majority of admissions, 94.1%, were either from acute units of the IRF's own facility or another facility. Prior to the event necessitating the inpatient rehabilitation stay, 95.3% of Medicare patients were living at home, 2.5% resided in assisted living facilities, and 2.0% were living in all other categories (i.e., board and care, subacute setting, chronic hospital) and 2% were missing this information. Among patients living at home, over half of all Medicare patients admitted to IRFs lived with family or relatives (58.6%), another third lived alone (34.7%), and 1.9% lived with friends, attendants or others.

Medical Information—The next section reports on selected variables from the Medical Information and Needs section of the IRF-PAI form. The breakdown for the 6 largest impairment groups is listed in ***Table 1.*** A substantial number of all Medicare IRF admissions were for orthopedic disorders, followed by stroke.

Table 1
Impairment group reported on admission as the percentage of total admissions

Impairment group	Percentage of total admissions
Orthopedic Disorders	41.5%
Stroke	17.1
Medically Complex	6.0
Cardiac	5.7
Neurologic Condition	4.6
Brain Dysfunction	3.4

Percentages may not add to 100 due to rounding.

The Impairment Groups Congenital Deformities and Developmental Disabilities were not reported, because the percentage was less than 1% of total admissions.

Medical Needs—This section of the IRF-PAI form consists of all voluntary items and examines patients' neurological status on admission and two items at admission and discharge—their swallowing status and whether the patient exhibits any clinical signs of dehydration.

Providers answered whether the Medicare patient was comatose or delirious 35% of the time on admission and when reported (Yes/No), they reported affirmatively just 1% of the time. Similarly, on admission or discharge, providers answered the items pertaining to the patient's swallowing status 36% and 35% of the time, respectively, and when reported, they reported approximately 7% and 6% of the time on admission and discharge that the patient was unable to ingest regular food. Finally, providers reported that only 1% of the time, either at admission or discharge, a patient demonstrated clinical signs of dehydration. For the majority of inpatient rehabilitation hospitalizations, 66% on admission and 67% on discharge, the dehydration item was not completed. This substantiated the TEP's concern that the neurological and dehydration

measures were not relevant to many IRF admissions, although they may be relevant for nursing facility stays. The data also showed that swallowing functions vary among rehabilitation patients, although only a small proportion of responses were given.[3]

Function Modifiers—The IRF-PAI section on Function Modifiers consists of ten items regarding bladder and bowel status as well as transfer and ambulation. This section is mandatory and was completed for all admissions. The next section of the form contains the 18-item FIM™ instrument. These items are mandatory items on admission and discharge but the goals column is voluntary. The goals are reported for each item for the inpatient rehabilitation stay. On admission and discharge, the individual FIM™ items were consistently completed, but the goal was missing on all FIM™ items, for an overall average of 49% of admissions. More hospitals reported these scores, but it was still less than half the admissions.

Discharge Information—The section on discharge information contains items on date of discharge, whether the Medicare patient was discharged against medical advice, whether there were any program interruptions, and information on where the patient was discharged. 74% of the Medicare admissions reported the patient were discharged to home, followed by discharge to a skilled nursing facility (9.4%), an acute facility (9.3%), assisted living facility (4.3%), subacute setting (2.3%), and intermediate or board and care (<1%). This is consistent with other similar analyses of IRF discharge destinations (Gage et al., 2004).

Quality Indicators—This section is comprised of 23 items, all completed on a voluntary basis and addressing topic areas pertaining to respiratory status, pain, pressure ulcers, and safety. All of these items were reported to the TEP. The first items, Shortness of Breath (SOB) on Exertion and at rest and Weak Cough/Difficulty Clearing Airway Secretions, were not completed for 75% of admissions. When completed, the items were answered in the affirmative for only 2% or less of all cases for all but the SOB on Exertion item. On admission, this item was answered "Yes" 7% of the time and at discharge, 5% of the time.

The item pertaining to pain addressed the highest level of pain reported by the patient on a scale of 0 (no pain) to 10 (worst pain possible) on admission and discharge. The item was not completed on admission 71.8% of the time and 72.1% at discharge. When completed, however, this item demonstrated variability among levels reported as well as improvements between admission and discharge (see ***Table 2***).

The items concerning pressure ulcers examined the highest current pressure ulcer stage, number of pressure ulcers, and PUSH Tool V. 3.0. The item addressing the highest current pressure ulcer stage was not completed for 77.7% of the Medicare admissions and when completed, Stage 0 (no pressure ulcer) represented 19.9% of admissions while Stage 1-5 were reported in only 2.3% of all admissions. On discharge, this item was not completed for 78% of admissions and when completed Stage 0 (no pressure ulcer) represented 20.1% of admissions while Stage 1-5 were reported in 1.9% of admissions. This suggests that among those reporting, few Medicare patients had pressure ulcers.

---

[3] It is unknown how many of the non-respondents would have had swallowing problems. A sensitivity analysis showed that response levels varied by facility, suggesting that non-responses were more likely to be related to facilities than to items.

Table 2
Level of pain reported on admission and discharge as a percentage of total admissions

Pain Level	Admission	Discharge
Missing	71.8%	72.1%
0	8.9	12.0
1-3	2.3	2.3
4-6	7.5	3.7
7-9	7.1	3.9
10	2.5	0.9

The number of pressure ulcers was not reported in 97.6% of admissions and 98.2% of discharges. Any number of pressure ulcers was reported for only 1.6% of admissions and 1.5% of discharges. The PUSH tool was not completed in 97% and 98% of admissions either on admission or discharge, respectively.

Further Analysis—After the presentation, the findings and next phase of the analysis were discussed with the TEP. The TEP proposed examining the data on both a national and facility level. On a national level, the range and standard distribution of selected IRF-PAI variables could be examined at admission and discharge and differences between these two timeframes could be analyzed by variable. Additionally, at the facility level, these same analyses could be undertaken to compare facilities by volume as well as by facility characteristics. The question remained whether there is sufficient volume and variation to examine meaningful differences on a facility level.

2.4.4 Second TEP Meeting Follow-Up

From the second TEP meeting, participants provided RTI with suggestions for several domains that should be added to the revised IRF-PAI and field tested. For the domains related to premorbid social networks, cognitive functioning, and swallowing, panel members suggested specific instruments or measures from which new items could be adapted. TEP members asked RTI to canvas the literature for items related to depression/mood and process measures and suggested several ways to construct an additional item to measure premorbid physical functioning. TEP members requested that RTI suggest one or more items for each of these domains and contact some of the panel members for input on which one(s) should be included before field testing the revised instrument.

Section 3 summarizes the literature review. Then Section 4, Primary Data Collection, presents the final set of items selected for the pilot test.

SECTION 3
LITERATURE REVIEW

RTI was asked to provide the TEP members with information on five additional domains that may be relevant quality indicators, but which are not currently included in the IRF-PAI. We conducted an extensive literature review and identified existing instruments that had already been validated in each of the five areas. This literature review presents the results of this search. In addition, items from the Rehabilitation Institute of Chicago's Functional Assessment Scales© (RIC-FAS) were also considered but are not included here in the literature review. A separate manual describes the range of scales available under the RIC-FAS (RIC, 1998). The five domains for further study are:

- Depression
- Pre-morbid social networks
- Pain
- Premorbid physical functioning
- Cognitive functioning.

In response to the TEP's suggestion, RTI conducted a literature review of the five domains identified by the panel. Each domain was assigned to an RTI staff member who was responsible for searching the literature, reading and analyzing the results, and identifying potential measures. To begin, key search terms were used for each domain. Searches were conducted using PubMed, Medline, and other relevant databases. In most cases searches were done in two steps: 1) a search of the literature for each domain (e.g., depression) and 2) a search of measures identified through step 1 (e.g., Geriatric Depression Scale). Relevant articles were entered into a database developed to capture key components of each measure. The data was then used to produce a literature review for discussion at the second TEP meeting, November 6 and 7, 2004. This review includes explanations of why each measure was examined.[4]

3.1 Depression

The TEP agreed that depression is important for risk adjustment because patients who are clinically depressed have poorer outcomes. Several studies have found that depression is a common psychiatric disorder of both institutionalized and non-institutionalized elderly. Although there is considerable disagreement in estimates of its prevalence (from 25% to 79%), there is general agreement that depression is associated with various negative outcomes (Gillen, Eberhardt, and Tennen, 1999).

Depression has been studied as an outcome measure, as well as a predictor of length of stay in a hospital or rehabilitation unit, rehabilitation efficiency, mortality, and as a risk adjuster. (Brink, 1982; Mossey, 1989; Hosking, 1996; Lyness, 1995; Davidson, 1994; Shinar, 1986;

[4] See Appendix F for the complete set of instruments in the Literature Review.

Arfken, 1999; Van Marwijk, 1995). However, the typical IRF length of stay is not long enough to see significant improvements in depression. Therefore, our review of possible measures focused on finding suitable items to use as risk adjusters.

A second issue that was considered in assessing the suitability of different measures was whether the population that the measure was developed for was similar to the Medicare inpatient rehabilitation population. Most of the existing depression rating scales have been developed and validated with younger, general population samples. These instruments tend to be heavily loaded toward measuring somatic items, such as loss of appetite and sleeplessness. While these symptoms can be indicators of depression in the elderly, they can also be side effects from medication, the result of physical deficits, or the aging process itself. Thus, such instruments may produce false positive results when used with the older populations.

When choosing instruments for review, we included a wide assortment so as not to leave out measures that may not be appropriate in their entirety, but instead might include a section or item of use. Instruments reviewed include:

- Center for Epidemiological Study of Depression Scale (CES-D)
- Hamilton Depression Rating Scale (HAM-D)
- Zung Self-Rating Depression Scale (SDS)
- Yale Depression Screen
- 4-item and 1-item Geriatric Depression Scale (GDS)

The CES-D, HAM-D, and SDS were all originally designed as general population measures. However, the CES-D has also been validated with a number of elderly populations, including stroke patients, the physically disabled, and the frail elderly. While most studies found the CES-D to be appropriate and valid, one study conducted with nursing home patients found that the CES-D took longer than other instruments to administer, and required the interviewer to frequently repeat several items. Ease of administration is a key factor in selecting an appropriate instrument. Instruments that involve several choices where subtle discrimination is required may be difficult to use with this population. Most researchers agree that a simple, easily understood instrument that is sensitive enough to distinguish between depression and other conditions with similar symptoms is essential for use with the geriatric population.

The HAM-D is a 21-item scale that was designed to be administered by a trained clinician. Administration is labor intensive in that the clinician must consider both the intensity and frequency of a symptom and then assign a rating value. The HAM-D is also commonly combined with a clinical interview to rule out endogenous depression. This makes the HAM-D inappropriate for the IRF-PAI.

The SDS is a screening instrument that provides information on current depressive symptoms. Although designed as a general population measure, it has also been tested with the elderly, including stroke patients and other physically ill and disabled populations. However, the potential for false positives and varying validity coefficients makes the SDS inappropriate for the IRF-PAI.

The Yale Depression Screen and the 1- and 4-item GDS were more promising because they were designed specifically for the geriatric population. The Yale Depression Screen is a 1-item yes/no question ("Do you often feel sad or depressed?"). It is typically used as a screening question to identify patients in need of further evaluation.

The 4-item GDS consists of four yes/no questions derived from the 30-item GDS. It is mainly used for assessment at admission, but has also been identified for use as a risk adjuster. Studies comparing the 4-item GDS to the 30-item GDS found that the 4-item measure has acceptable levels of construct validity. The 1-item version of the GDS ("Do you feel that your life is empty?") has not been well tested yet, but has shown promising signs as a quick and simple method of identifying elderly patients with depressive symptoms that may require a follow-up evaluation.

After field testing both the Yale Depression Screen and the 4-item GDS, the psychologists on the TEP recommended using the Yale Depression Screen. The YDS has been tested with the elderly in outpatient and inpatient settings (Mahoney, 1994; Watkins, 2001). A study of the YDS in patients recovering from strokes found it to be helpful for clinicians in screening for depression after stroke. In this study the values for the YDS were sensitivity 86%, specificity 78%, positive predictive value 82%, negative predictive value 82%, and 82% of the cases were classified correctly (Watkins, 2001).

After reviewing these instruments, several TEP members suggested that in addition to measuring mood/depression, a measure of engagement in treatment would be beneficial. At the TEP's suggestion, we reviewed the Rehabilitation Institute of Chicago Functional Assessment Scale (RIC-FAS). Subsequently, both a mood/depressions scale and an engagement scale from the RIC-FAS were added to the IRF-PAI for testing. The items use a 7-point rating scale to measure a patient's motivation to actively participate in treatment.

3.2 Premorbid Social Network

At the first TEP meeting, the panel agreed that premorbid functioning has an important effect on rehabilitation outcomes and should be used as a risk adjuster. A search of the literature found convincing support for the importance of social involvement and emotional support as independent predictors of outcomes. The literature review focused on family functioning and social integration to tap into these areas. The following instruments were identified for further examination:

- McMaster Family Assessment Device (FAD)
- Lubben Social Network Scale (LSNS)
- Debilitating Beliefs Scale (DBS)

The McMaster Family Assessment Device (FAD) is a 60-item measure, but a subscale of general family functioning (GF) has been tested in a population of elderly stroke survivors to see whether family functioning improves adherence to treatment protocol. The FAD is a paper and pencil survey that is filled out by a patient's family. It has been used to predict outcomes and for risk assessment to determine how capable a person may be of getting help for ADL after discharge.

The Lubben Social Network Scale (LSNS) is a 10-item scale that measures the nature of social networks, including number, frequency, closeness, and living arrangements. The measure was designed and tested on an elderly population and can be used to predict outcomes and for risk assessment to determine a patient's capability of getting help for activities of daily living (ADLs) or IADLs after discharge.

The Debilitating Beliefs Scale (DBS) was chosen for review because emotional reactivity and debilitating beliefs have been identified as having a strong negative relationship with rehabilitation outcomes (Melamed, 1999) The DBS is an 11-item scale designed to identify personality and cognitive variables that predict poor adjustment following myocardial infarction. This instrument is not suitable for cognitively impaired patients.

The panel members agreed that a measure of social support should be included in the IRF-PAI instrument as a risk adjustor. The TEP recommended a review of the literature to identify items that could be used to measure premorbid social networks. Subsequently, items adapted from the Outcome and Assessment Information Set (OASIS) were presented to the TEP and added to the IRF-PAI instrument.

3.3 Pain

At the first TEP meeting the panel discussed measuring pain at both admission and discharge. At admission, the TEP suggested taking a measure of pain severity to apply as a risk adjuster. At discharge, they suggested measuring the degree to which pain interferes with activity for use as an indicator or outcome.

The personal nature of pain makes it difficult to assess in a standardized fashion. Pain cannot be directly observed by clinicians. As a result, assessment is primarily dependent on patient self-report. This presents difficulties in measurement because individuals may perceive pain differently (Jensen and Karoly, 1992). In addition, self-report pain measures may be more difficult for cognitively impaired older adults. To complicate matters further, it has been demonstrated that pain is associated with psychosocial factors and depression in the elderly (AGS Panel on Chronic Pain in Older Persons, 1998).

In searching the literature for instruments meeting the requirements specified by the TEP we found that the literature lacks an integrated overview of pain assessment techniques and critical evaluation of the methods commonly used. To narrow the list of potential instruments we excluded palliative pain measures in favor of instruments that assess chronic, neuropathic, musculoskeletal, or arthritic pain (i.e., types of pain that are likely to be encountered in a rehabilitation setting). Measures of pain specific to certain body parts (e.g., back, head, neck, knee) were also excluded because their focus was too narrow and not necessarily generalizable to other types of pain. The following instruments were selected for further consideration:

- Numerical Rating Scale (NRS)
- Geriatric Pain Measure (GPM)
- Minimum Data Set (MDS)
- Faces Pain Scale (FPS)
- Musculoskeletal form of the Medical Rehabilitation Follow-Along (MRFA)
- McGill Pain Experience Instrument (CPEI)

The Numerical Rating Scale (NRS) is a measure of pain intensity expressed by a numerical scale. On a scale of 0 to 10, 2 or 3 is mild pain, 5 moderate pain, and 7 or higher severe pain. The scale has also been tested as a 5-point and 100-point scale. The NRS is one of the most commonly used scales in hospitals and other medical settings. The IRF-PAI includes a 10-point NRS.

The Geriatric Pain Measure (GPM) consists of 24 items (22 dichotomous yes/no questions and 2 items scored categorically on a 0 to 10 scale). The GPM can be used as an assessment of pain intensity, pain-related functional status, mood, and quality of life. The GPM provides more information that may be pertinent to rehabilitation than the unidimensional scales that measure only pain intensity. While the GPM includes an 11-point Numerical Rating Scale measuring current pain that is very similar to the current inpatient rehabilitation services pain measure, it includes additional information on pain not currently measured. This additional information includes the effect of pain on function, mood, engagement in activities, and quality of life. Functional status and how it is impacted by pain is particularly important among older people. Rehabilitation can be delayed, there can be an increased need for care, and physical function as well as quality of life can be reduced because of pain.

The Minimum Data Set (MDS) is part of the federally mandated process for clinical assessment of residents in Medicare and Medicaid certified nursing homes. Licensed health care professionals complete assessments to measure residents' functional capabilities and to identify health problems. The MDS includes a pain quality measure to identify residents who are experiencing pain. The MDS pain measure consists of 2 items assessed on a 3-point Likert scale with verbal descriptors.

The Faces Pain Scale (FPS) consists of visual depictions of faces representing increasing levels of pain intensity across a 7-face continuum with a range of 1-7. The FPS was originally developed for children, but several studies have found it works well with adults, especially the elderly.

The Musculoskeletal Form of the Medical Rehabilitation Follow Along (MRFA) measures quality of daily living, including physical function, pain, satisfaction, and emotional/psychological well-being.

The McGill Pain Questionnaire (MPQ) measure consists primarily of three major classes of word descriptors—sensory, affective, and evaluative. Patients use the word descriptors to specify subjective pain experience. The measure also includes an intensity scale and other items to determine the properties of pain experience.

The Chronic Pain Experience Instrument (CPEI) was designed to measure persistent, non-malignant pain. The measure consists of 24 items, all using a visual analogue scaling. The literature review did not include specific information about the pain items.

At the second TEP meeting the panel agreed that the item currently on the IRF-PAI to measure pain (a 0-10 rating of the highest level of pain reported by the patient during the assessment period) should remain as a mandatory item and be used for risk adjustment. This measure is used by most hospitals, so moving to a different item would create additional burden for the hospitals. The TEP recommended further research to determine whether the current item is useful for risk adjustment for the FIM™ change score and suggested another search of the literature for studies that validate the use of pain as a risk adjuster in the rehabilitation setting. Subsequently, a process measure on the methods of pain management used during the stay and an assessment of whether pain limits the patient's ability to participate in the rehabilitation process were added to the IRF-PAI.

3.4 Physical Functioning

At the first TEP meeting, members agreed that a measure of physical functioning is crucial in evaluating care provided by IRFs. The Functional Independence Measure (FIM™) is the current standard measure of physical functioning in rehabilitation settings. It was developed and tested by rehabilitation clinician specialists specifically for expressing the needs of patients treated in the IRF setting. The FIM™ measures impairment levels in 18 activities including a set of self-care, sphincter control, transfers, locomotion, communication, and social cognition items. Each item has a 7-point scale to measure the level of assistance needed from complete independence (7) to total assistance (1).[5] The 7 point scale allows for significant variation in results and is a good predictor of resource use and burden of care. This is particularly important for those individuals with more severe physical functioning limitations; ceiling effects are less likely and therefore identification of more extreme deficits and improvements are more likely using this scale.

Despite the dominance of the FIM™ in rehabilitation settings, a review of the literature was conducted to identify other potential measures of physical functioning. Four measures that were most frequently mentioned and used in studies were chosen for further examination:

- Medical Outcomes Study Short Form 36 (SF-36)
- Sickness Impact Profile (SIP)

[5] A "0" code was added to the IRF-PAI version of the tool to denote at admission whether the type of activity occurs.

- Nottingham Health Profile (NHP)
- Extended Activities of Daily Living (EADL)

The Medical Outcomes Study Short Form 36 (SF-36) is a measure of perceived health status. The measure includes 10 questions related to physical functioning, which would supplement much of the motor portion of the FIM™ as included in the IRF-PAI. The SF-36 has been used as a needs assessment, risk adjustor, and outcomes tool. The literature suggests that it can be used to measure change in outcomes before and after rehabilitation. The SF-36 has been widely tested with the elderly and those with physical impairments, but there is some evidence to suggest that validity and reliability of the SF-36 is low for the cognitively impaired. There is also some evidence that suggests that the questions and scaling of the instrument do not permit much differentiation of physical limitations, particularly among the severely disabled. Further, it is based on self-report while the FIM™ typically is based on observed or tested functional levels.

The Sickness Impact Profile (SIP) is a generic measure of health status, including physical functioning. The measure consists of 136 items requiring respondents to identify illness behaviors. Scores from 12 subscales are combined into the Physical Dimension and Psychosocial Dimensions. The SIP has been used to measure outcomes of care and individual patient progress. Like the SF-36, the SIP would supplement much of the motor portion of the FIM™ as included in the current IRF-PAI.

The Nottingham Health Profile (NHP) was designed to measure generic health status and well-being. It consists of six domains, including physical mobility. The measure consists of 38 statement requiring “yes” or “no” answers. The NHP was developed on a general population, but studies suggest it has been used successfully with individuals with disabilities, although there are floor effects and item non-response with severely disabled individuals. Like the other physical functioning measures discussed, the NHP would supplant much of the motor portion of the FIM™ as included in the current IRF-PAI. However, because of the NHP’s “yes” or “no” response format, it would not be as sensitive to small differences in individuals’ physical functioning as the current IRF-PAI physical functioning items.

The Extended Activities of Daily Living (EADL) is widely used as a measure of disability and has been well tested with the elderly. The measure consists of 22 items thought to be important for daily living at home. The items are grouped into four categories, one being mobility. Like the other measures reviewed, the EADL would supplant much of the motor portion of the FIM™ as included in the current IRF-PAI. However, there is some evidence that some of the items in the EADL are not relevant to the disabled because they focus on independent living, and therefore the EADL may not be effective at differentiating among the severely disabled.

At the second TEP meeting members agreed that the level of physical functioning pre-injury is important for risk adjusting a change in Functional Independence Measure (FIM™) scores and should be included in the IRF-PAI instrument. The TEP considered several scales that measure physical functioning, but decided that any scale used to measure physical functioning should be comparable to scores obtained by the FIM™ so data can be compared at admission

and discharge. TEP members agreed that the best solution would be to replace the "Goal" checkboxes with "3 months prior to onset" so the current FIM™ items could be used to collect a measure of premorbid functioning.

3.5 Cognitive Functioning

At the first TEP meeting members agreed that cognitive measures are needed for risk adjustment of outcomes, and that the possibility of adding items to the IRF-PAI could improve its sensitivity to the cognitive correlates of functional outcomes. As the IRF-PAI already contains the FIM™, which includes some cognitive measures, any additional items would have to predict outcomes more efficiently than the FIM™ cognitive items.

A review of cognitive functioning measures was conducted by reviewing literature and talking to experts in the field. Although some studies found no significant relationship between cognition and recovery of function, numerous other researchers report that cognitive status does affect function (Resnick and Daly, 1997). Similarly, there was little consensus among experts regarding the use of cognitive assessment instruments. One person contacted explained that the test chosen depends on the individual patients and that a wide variety of instruments are used on a regular basis, all of them modified as necessary. Based on suggestions from experts and a review of seminal studies, we chose the following measures for further evaluation:

- Mini-Mental State Examination (MMSE)
- Functional Independence Measure–Cognitive items N-R (FIM™)
- Cognitive Capacity Screening Exam (CCSE)
- Lowenstein Occupational Therapy Cognitive Assessment (LOTCA)
- Direct Assessment of Functional Status Scale (DAFS)
- Executive Interview (EXIT)
- Clock Drawing Task 1 and 2 (CLOX)
- Cognitive Impairment Diagnosing Instrument (CIDI)
- Cognistat/Neurobehavioral Cognitive Status Exam
- Assessment of Language Related Functional Activities (ALFA)
- Ross Information Processing Assessment-Geriatric (RIPA-G)

The Mini-Mental State Examination (MMSE) is used in screening for cognitive impairment severity and as serial documentation of cognitive change. The MMSE consists of 11 open-ended and performance-based items. The items are grouped into seven categories of cognitive function: Orientation to time, Orientation to place, registration of three words,

attention and calculation, Recall of three words, Language, and Visual construction. Although the MMSE is widely used, both sensitivity and specificity are affected by number of years of education. Scores have also been shown to be affected by social class, socioeconomic status, and age.

The Functional Independence Measure (FIM™) assesses independence across motor and cognitive domains through therapists' direct observations of performance. It consists of 18 items, each scored on a scale of 1-7, the highest score reflecting complete independence (Granger, et al., 1995). Cognitive domain items include comprehension and expression (communication), social interaction, problem-solving, and memory (social cognition). The motor and cognitive domains are nested within the global concept of burden of care that is measured by the total FIM™ and are more appropriate than the total FIM™ score for answering more clinically focused questions about general types of disabilities (Stineman, et al., 1996). Several studies find the FIM™ to be a good generic indicator of disability. While typically performance based, a version of the FIM™ is also administered by phone.

Cognitive Capacity Screening Exam (CCSE) is a 30-item measure designed to "diagnose diffuse organic mental syndromes in nonpsychiatric patients." The content areas measured include orientation, memory, attention, calculations, and concept formation with scores ranging from 0-30 (scores of less than 20 are indicative of cognitive impairment. A TEP member (Christine Baron: Co-director of the Stroke & Recovery Program, and manager in Speech & Pathology Service at National Rehab Hospital) indicated that the CCSE is used regularly in the rehabilitation field.

The Lowenstein Occupational Therapy Cognitive Assessment (LOTCA) consists of 20 subtests across four areas of orientation, perception, visual motor organization, and thinking operations. While the LOTCA may be slightly superior to the FIM™ cognitive scale and the MMSE as it relates to functional outcome parameters, it is lengthy and burdensome to administer.

The Direct Assessment of Functional Status Scale (DAFS) assesses seven functional areas, with specific tasks associated with each area: time orientation, communication, transportation, financial, shopping, grooming, and eating. The measure includes a total of 106 items, which are administered by in-person interviews with direct observation of performance. In an effort to increase its sensitivity to early dementia and otherwise impaired patients, the DAFS is currently being revised to include medication management, food preparation, and taking telephone messages. One of the TEP panel members (Peter Lichtenberg: Director of the Institute of Gerontology and associate professor of psychology at Wayne State University) recommended the DAFS as a good instrument from which to choose some possible cognitive items for the IRF-PAI.

The Executive Interview (EXIT) was designed to measure executive cognitive function among the elderly. The EXIT is an in-person, performance-based interview consisting of 25 items, scored 0-48, with higher scores indicating greater executive dyscontrol. The EXIT is simple and has clinical face validity because many of the items are derived from routine clinical procedures. (Royall, 2002).

The Clock Drawing Task 1 and 2 (CLOX) was designed to measure executive impairment in the elderly. The first part of the exam (CLOX1) involves drawing a clock that says 1:43. The CLOX 2 involves the examiner drawing the 1:43 clock and having the subject copy it. CLOX scores have been found to be strongly correlated with cognitive impairments as measured by the EXIT and the MMSE (Royall, 1998).

The Cognitive Impairment Diagnosing Instrument (CIDI) is designed to measure cognitive function in the elderly. It includes 73 items across 10 subscales (i.e., short-term memory, long-term memory, orientation to time, orientation to place, memory registration, concentration/calculation, judgment, object naming, abstract thinking, and higher cortical functions). The data are gathered through in-person semi-structured interviews averaging 20 to 52 minutes to complete (the longer time is for one study conducted with a Korean speaking sample).

Cognistat/Neurobehavioral Cognitive Status Exam has been used to determine cognitive status among individuals with traumatic brain injury to determine readiness for rehabilitation, cognitive functioning, and cognitive impairment among psychiatric patients. The measure consists of 11 subtests assessing consciousness, attention, and orientation, language construction, memory, calculation, and reasoning. Subjects who pass the screening items are considered to be cognitively intact in that domain. Two TEP members (Dr. Eliot Roth of the Rehabilitation Institute of Chicago and Dr. Bruce Gans of the Kessler Rehabilitation Corporation) stated that the Cognistat/NCSE is commonly used in the field to assess cognitive function.

The Assessment of Language Related Functional Activities (ALFA) is a measure of functional skills on a set of language-related tasks. The ALFA is designed for persons between the ages of 16 and 91 who can understand the directions of subtests, who are able to formulate the necessary responses, and who have some familiarity with the functional areas assessed. It consists of 10 subtests: telling time, counting money, addressing an envelope, solving daily math problems, writing a check and balancing a checkbook, understanding medicine labels, using a calendar, reading instructions, using a telephone, and writing a phone message. The total time for administration can range from 30 minutes to 2 hours (some subtests are timed and others allow for as much time as needed). A TEP member (Christine Baron: Co-director of the Stroke & Recovery Program, and manager in Speech & Pathology Service at National Rehab Hospital) indicated that the ALFA is regularly used in the field and is, in fact, a favorite because the norms are more applicable and it is more functionally based.

The Ross Information Processing Assessment-Geriatric (RIPA-G) is designed to identify, describe, and qualify cognitive-linguistic deficits in the geriatric population following traumatic brain injury. The battery consists of 10 subtests, 2 supplemental subtests, and a record form with 3 subsections to include background, medical, RIPA-G test, and retest information. It measures various cognitive or linguistic processes, including memory, orientation, organization, problem solving, auditory processing, knowledge of general information, reading, and word finding. A TEP member (Christine Baron: Co-director of the Stroke & Recovery Program, and manager in Speech & Pathology Service at National Rehab Hospital) indicated that this instrument is regularly used in the rehabilitation field. She also noted, however, that it is not well-liked because the norms do not always hold true and it is not functionally based.

At the second TEP meeting the panel discussed the different aspects of cognition, how to best assess cognitive functioning, and which instrument(s) to use. Because of its complexity, the TEP members did not agree which instrument to recommend. After further discussion with the TEP, it was decided that the FIM™-based cognitive functioning measures already on the IRF-PAI should remain in use.

SECTION 4
PRIMARY DATA COLLECTION

4.1 Purpose

Based on the suggestions of the TEP members, CMS directed RTI to conduct a pilot test of the proposed new quality-related items that might be included in a revised IRF-PAI instrument. By collecting primary data on these potential new items, RTI was able to create a dataset to evaluate the relationships between the proposed new quality items and potential IRF outcome measures. Second, the process provided an opportunity to test the feasibility of using each item in different types of inpatient rehabilitation hospitals and with different populations. Following the advice of the TEP, the IRF-PAI instrument was revised to add a set of items on pre-morbid social networks, pre-morbid physical and cognitive functioning, depression/mood/engagement, pain, and pressure ulcers.[6] Most of these items were selected from the literature review presented in Section 3; however, a few of the items, such as the pre-morbid functioning items, were developed to meet the recommendations of the TEP. The revised IRF-PAI instrument and explanations of recommended changes to the IRF-PAI are displayed on pages 10 and 15 of the IRF-PAI coordinator manual (see ***Appendix G***).[7]

This section describes the primary data collection effort, including the site selection process, the IRF-PAI revisions, and the results of this effort, including descriptive statistics on each of the new items and comments from the IRFs that participated in the pilot study regarding the strengths or challenges associated with each item.

4.2 Site Selection

Based on direction from CMS, RTI selected 11 hospitals from a list of all Medicare-participating IRFs. The list was generated from the IRF-PAI data and subset to identify IRFs with at least 300 admissions per year and who answered 30% or more of the voluntary items. These criteria were used to ensure that an IRF would have enough cases in one month to provide a reasonable sample size, and that volunteer IRFs would be better able to participate without a major disruption in their units. Hence, it was felt that hospitals that attempted to complete a small proportion of the voluntary items would be better able to respond to the request for pilot site participation.

In addition to unit size and IRF-PAI experience, IRFs were also selected for variation in geographic location, ownership, and whether they represented freestanding facilities or rehabilitation units within a larger acute hospital. The final selection included 9 hospitals

---

[6] The PUSH tool was originally going to be excluded, but input from wound care nurses at the Site Coordinator training pointed out that changes in an ulcer could not be documented without that level of detail. Discussion with the hospitals suggested that most were collecting this type of information and either recording it with scores or photographs in the patient record; it just was not submitted to CMS. After discussion with the CMS team, this measure was retained in the pilot test.

[7] The pilot test included several existing IRF-PAI items that were voluntary (3 point swallowing item [28B], 10 point pain item [50A], and the pressure ulcer items [51A-F]). All new items have an ‘N” in front of the number.

geographically dispersed across the country. About half the IRFs were freestanding and half were units. While this proportion did not represent all IRFs, it did ensure an adequate sample size of 9 IRFs. "Ownership" selection criteria included a mix of proprietary, non-profit, and government-owned facilities, including at least one hospital chain.

4.3 Data Collection Methods

Each participating IRF was asked to appoint a study coordinator to work with RTI and manage the data collection effort, including training staff. To ensure that the study coordinators received consistent training on the revised IRF-PAI forms, RTI staff conducted a one day "train the trainer" session for the study coordinators in the RTI offices in Waltham, MA on September 28, 2004. The training session was also attended by CMS staff, including the project officer, Rita Shapiro. Training was provided by RTI staff with the assistance of Anne Deutsch, a TEP member with experience training hospital staff on the use of the Uniform Data System for Medical Rehabilitation. The day consisted of an overview of the study, a review of the PAI changes on the pilot form, practice vignettes to explain the swallowing and mood/depression scoring scales, an overview of the data collection process and data management, an initial discussion of site visits, and time for questions and answers (see ***Appendix G***).

During the discussion of the data collection process, site coordinators were asked to consider the range of disciplines of staff members who would be involved in completing the various items. These included: occupational therapists, speech therapists, physical therapists, social workers, case workers, nurses, psychiatrists, and possibly psychologists or psychiatrists. As with the IRF-PAI, each coordinator was asked to select the most appropriate staff discipline for each section of the form. In the case of a few IRFs, the study coordinator or RTI staff used the IRF-PAI form to create discipline-specific worksheets for the respective disciplines. This was consistent with their current IRF-PAI practices.

Between October 10 and October 26, 2004, teams of 2 RTI staff conducted two-day site visits at the nine participating IRFs. The first day included interviews with the rehabilitation management team and a meeting with the study coordinator. The meeting with the hospital management team consisted of the medical directors of rehabilitation, quality, nursing, therapy, and case management. The purpose was to introduce the study, answer any questions, and discuss issues involved in the IRF-PAI administration. The meeting with the study coordinator and, if applicable, the other staff involved in managing the IRF-PAI data entry, was to discuss the study logistics and address any questions about study administration or staff trainings. The current roles of each IRF's PPS coordinator ranged from collecting the FIMS and related IRF-PAI data from patient records and using it to complete all patient IRF-PAIs at a hospital (but not treating the patients), to managing the unit staffs' completion of segments of the IRF-PAI form and using that to complete the form. A few hospitals used the coordinator to answer staff questions and manage the data submission to CMS, but the forms were completed by the unit staff. Each IRF was asked to use their usual method for collecting the pilot data so that it was consistent with their respective IRF-PAI data collection process.

The second day of the site visit was spent attending the training sessions. Each session was one hour in length; coordinators organized 1-6 sessions to train staff on each shift and in each relevant discipline. RTI staff introduced the study and the opportunity it presented to

provide the federal government with feedback on the proposed IRF-PAI revisions. The coordinators trained the unit staff on the use of the revised IRF-PAI form. Training was also provided to the weekend staff, and coordinators were asked to be present and available on the units during the first several days of data collection to answer questions.

Data collection began on the day following the site visit or on another agreed-upon date.[8] The revised IRF-PAI was completed for each Medicare admission on or after the start date. Discharge information was also collected for the admission cohort when they were discharged. The target number of admissions for each hospital varied by IRF, depending on the size of the unit(s), average number of Medicare admissions, and length of stay. RTI's goal was to collect a sample size of at least 500 cases.

RTI staff provided each coordinator with the materials needed to conduct the pilot test, including the staff training manuals and copies of the revised forms. The forms were color coded to distinguish between admission and discharge. This made it easy to identify whether each patient had both forms completed. Each coordinator was given a notebook with a copy of the coordinator's manual, tracking sheets, a feedback form, and pre-paid FedEx labels. The tracking sheet was designed for coordinators to list study patients' names, Medicare health insurance numbers, admission and discharge dates, and FedEx dates for both the admission and discharge data to track when the information was shipped to RTI.

In addition to the revised IRF-PAI tools, each Coordinator was given a set of feedback sheets for use by the Coordinator and units' staff to document the strengths and weaknesses of the revised form and offer suggestions for other improvements. This was especially emphasized with the swallowing and depression items, as multiple options for potential items were included on the revised IRF-PAI form. RTI was particularly interested in feedback on these items, which would inform CMS decisions on which, if any, of the item options would be included in a potential final revised IRF-PAI instrument.

During the 6-8 week period of data collection, the Coordinators mailed weekly packages to RTI with the most recent admission or discharge forms, tracking sheets, and feedback forms. One hospital opted to enter the information into an access database provided by RTI in addition to submitting paper versions of the forms. RTI staff members were available to answer questions and provide technical assistance to the study coordinators throughout this period.

RTI also established a list-serve with all the email addresses of the IRF study coordinators, Anne Deutsch, and RTI staff. The list-serve enabled RTI and the coordinators to have an ongoing conversation, with the benefit of immediate problem solving and information sharing.

[8] Data collection did not usually begin on the weekend since staffing is different then, so some data collection periods began on the Monday following the site visit instead of the next day.

4.3.1 Database Development

The submitted data were entered into an electronic file created in Microsoft Access. This allowed the entry page to look like the form. The database contains one table for each page of the revised patient assessment instrument in addition to a master form. This master form served primarily as a user interface for data entry and cleaning. One facility sent a completed database through secure file transfer protocol, and RTI created a separate Access file for each of the remaining eight hospitals. These files used restricted variables so that only answers allowed by the pilot test-revised IRF-PAI could be entered into the database. This helped to ensure accuracy of data entry and to determine data-related questions to be resolved with IRFs. After calling the facilities to resolve data issues, we entered the appropriate revisions into the Access files before reading these files into a SAS program for data checking and analysis.

We used the SAS system to verify all entries and to link the relational files into a single analytic file. This required linking the five pages per patient into one case record. Records were linked using Medicare ID numbers for the beneficiary, the provider, and their admission date which was included on each of the five tables corresponding to each page of the patient assessment instrument. We then combined the files for each of the nine IRFs into a single analytic file. Upon completing this analytic file, we created new variables to re-calculate the total scores for the FIM PUSH tool and the geriatric depression scale. These calculated variables were used instead of the hand-recorded items in the analysis as they corrected for human error.

4.4 Pilot Test Instrument

The revised IRF-PAI tool used in the pilot test is provided in ***Appendix G (***pages 15-19**).** Changes to the original form are summarized in the training manual. The changes were based on the recommendations of the TEP with CMS approval. Many of the pre-morbid social network items were taken from the OASIS form used in Medicare's home health benefit, a benefit providing services to similar populations who are at the next level of care—at home but needing some additional therapeutic support to be successful in their rehabilitation. Several items were selected from other existing instruments, such as the Rehabilitation Institute of Chicago's RIC-FAS instrument. Still others, such as the pre-morbid functional level, resulted from discussions with CMS personnel and one-on-one conversations with several TEP members. RTI staff reviewed the list of possible additional items or modifications and made an initial set of revisions to the IRF-PAI. In some cases, small changes were made to the wording of the item or possible responses were added or modified. In addition, some items were entirely removed or replaced with items the TEP felt were more appropriate for the IRF population. Several new items were added to address domains not previously included on the IRF-PAI. RTI staff summarized these revisions and TEP members were asked to review and comment on each item that was modified, added, or deleted. TEP members' responses were incorporated into the final revision of the IRF-PAI used during the pilot test. The form was pre-tested in one facility and slight revisions resulted in its content.

The following subsection provides a brief summary of the revisions made to the IRF-PAI and the rationale behind these revisions. We discuss descriptive statistics for each revised indicator to understand how the indicators performed in the field and review comments made by site coordinators and other field staff about the administration of the indicator.

4.4.1 Pre-Morbid Social Network: Changes to the Admission Information

Several changes were made to the "Admission Information" collected on the original IRF-PAI. First, responses 4 and 5 for item 15 were revised to read "Intermediate Care/LTC Facility" and "Skilled Nursing Facility," respectively. This distinguished between patients residing in long-term care settings versus those being admitted from a short-term post-acute or skilled setting. The rehabilitation potential for the two groups would be different and this will better distinguish between the two. In addition, items 16 through 20 on the IRF-PAI were replaced with a new set of items (N16 through N21 on the revised IRF-PAI) adapted from the Living Arrangement and Supportive Assistance items (M0300 through M0380) on the Outcome and Assessment Information SET (OASIS-B1). These new items addressed TEP members' request that the IRF-PAI include questions assessing premorbid social support, since the amount of support available affects the probability of a patient being discharged back into the community. Using the OASIS items also provided consistency between Medicare services in measuring similar concepts. Therefore, the TEP recommended this item be included for risk-adjustment purposes in the IRF-PAI instrument.

For item N16 on the revised IRF-PAI, IRF staff was asked to mark where a Medicare patient resided prior to admission (***Table 3***). Nearly 96% of the patients in our sample reported living in their owned/rented residence or that of a family member (88.7% and 7.2%, respectively). The completion rate for this item was extremely high, with only one missing response.

Table 3 (N16)
Prior to hospitalization: Pre-admission residence

Residence	Percent
1 - Patient's owned/rented residence	88.7
2 - Family member's residence	7.2
3 - Boarding home/rented room	1.0
4 - Board and care or assisted living facility	2.0
5 - Long-term care facility	0.8
6 - Other	0.4

Sample size = 514

Frequency missing = 1

Percentages may not add to 100 due to rounding error.

Item N17 assesses who the patient is living with prior to admission and responses included alone, a spouse/significant other, family member, friend, or other paid help (***Table 4***). A majority of patients reported living with another person prior to admission: 51.7% reported living with a spouse/significant other and 11.1% reported living with another family member. One-third of the sample population (33.9%) reported living alone. Only two patients reported living with a friend. The completion rate for this item was also very high, with only two missing responses.

Table 4 (N17)
Prior to hospitalization: Patient lives with

Lives with	Percent
1 – Lives alone	33.9
2 – With spouse or significant other	51.7
3 – With other family member	11.1
4 – With a friend	0.4
5 – With paid help	2.1
6 – Other	0.8

Sample size = 513

Frequency missing = 2

Percentages may not add to 100 due to rounding error.

Item N18 identifies the individuals who provide any type of assistance to the patient, including relatives, spouse or friends, and paid help (***Table 5***). Patients could report more than one response. Responses for this item were fairly well distributed, a desired trait in item response. 38.6% of the respondents reported that no care was needed prior to the acute event, while 27.1% of patients reported assistance from relatives, friends or neighbors; 33.2% from a person residing in the home; and 18.8% from paid help. In seven cases, IRF staff was unable to assess what type of assistance the patient received.

Table 5 (N18)
Prior to hospitalization: Assisting person(s) (check all that apply)

Assisting persons	Percent
0 - No care needed	38.6
1 - Relatives, friends or neighbors living outside home	27.2
2 - Person residing in the home	33.2
3 - Paid help	18.8
4 - None of the above	0.8
5 – Unknown	1.4

Sample size = 515

Categories are not mutually exclusive.

For item N19, IRF staff was to assess the categories of assistance provided by the primary caregiver identified in item N18 in the 3 months prior to onset (***Table 6***). Responses ranged from ADL and IADL assistance to less intense forms of support such as environmental and psychosocial support or services of a financial or health care agent. Responses for this item were well distributed across most categories. 51.3% of respondents reported requiring IADL

assistance and 21.8% required ADL assistance. 37.3% of patients required some type of environmental support and 42.9% received psychosocial support. Nearly 20% of caregivers facilitated the patient's participation in medical care. In 16% of the cases, staff were unable to assess what type of primary caregiver assistance was received.

Table 6 (N19)
Prior to hospitalization: Type of primary caregiver assistance in the 3 months prior to the onset (check all that apply)

Type of assistance	Percent
1 - ADL assistance	21.0
2 IADL assistance	51.3
3 - Environmental support	37.3
4 - Psychosocial support	72.9
5 - Advocates or facilities patient's participation in appropriate medical care	18.8
6 - Financial agent, power of attorney, conservator of finance	7.6
7 - Health care agent, conservator of person, or medical power of attorney	10.7
8 Unknown	16.7

Sample size = 515

Categories are not mutually exclusive.

Item N20 was answered at admission and discharge and identified the person who is likely to be "in charge" of providing and coordinating the patient's care once the patient is discharged (***Table 7***). The admission item provides information on the patient's pre-admission social support network and the discharge item identifies the patient's actual support network. At admission, 30.9% of patients reported they would be responsible for their own care while over 55% felt their spouse or son/daughter would take the lead in providing care (33.2% and 24.9%, respectively). Approximately 9% of respondents felt that other family members (3.7%), paid help (2.9%), or more than one individual (2.1%) would be responsible for care.

Responses change somewhat at discharge. The proportion of individuals reporting they would be responsible for their own care dropped to from 30.8% to 14.4% while the proportion reporting paid help increased from 2.9% to 12.2%. A higher proportion of patients also responded that more than one person would provide care (7.2%, up from 2.1%). The proportion that responded at discharge that a spouse or son/daughter would provide care stayed relatively similar (30.9% and 21.9%, respectively). Notably, while only 1.4% of patients responded that they did not know who would take the lead in providing or managing care at the time of admission, this proportion increased to 8.0% at discharge.

Table 7 (N20)
Discharge expectations at admission and discharge: Primary caregiver likely to take lead responsibility for providing or managing the patient's care

	Percent	
Caregiver	Admission	Discharge
0 - Self	30.9	14.4
1 - Spouse or significant other	33.2	30.9
2 - Daughter or son	24.9	21.9
3 - Other family member	3.7	3.5
4 - Friend/neighbor/community/church member	1.0	1.9
5 - Paid help	2.9	12.2
6 - More than one person	2.1	7.2
7 - Unknown	1.4	8.0

Sample size = 515

Percentages may not add to 100 due to rounding error.

Item N21 was also answered at admission and discharge and identified the frequency of all help available from the primary caregiver identified in item N20 (***Table 8***). The item was collected at discharge because initial expectations may change as the patient's caregiver develops a better understanding of what may be involved in care giving. At admission, 55.0% of patients reported their caregiver could provide care several times during the day and night, while 11.3% could provide care several times during the day. About 21% of patients responded "unknown" to this question. While responses were relatively similar between admission and discharge, the proportion of patients responding "unknown" decreased at discharge to 17.1%, from 21.0% at admission. Responses for several times during the day and night and several times during the day increased slightly at discharge (57.3% and 13.2%, respectively).

Table 8 (N21)
Discharge expectations at admission and discharge: How often could patient receive assistance from the primary caregiver (after discharge)

	Percent	
Caregiver assistance	Admission	Discharge
1 - Several times during day and night	55.0	57.3
2 - Several times during day	11.3	13.2
3 - Once daily	7.0	6.2
4 - 3 or more times a week	2.3	3.7
5 - 1 to 2 times per week	1.9	1.2
6 - Less often than weekly	1.6	1.4
7 - Unknown	21.0	17.1

Sample size = 515

Percentages may not add to 100 due to rounding error.

Staff from 4 of the 9 IRF sites reported some difficulty answering N19. If the answer to N18 was "no care needed," staff members were unclear whether they should leave N19 blank or answer the question anyway. Several suggestions were offered for how to revise this item to avoid confusion: two facilities suggested adding a "none" category to N19; one facility suggested adding a "not applicable" response, and one facility recommended N18 could have a skip statement, so if N18 was marked as "no care needed" item N19 would be skipped.

Some problems were reported for items N20 and N21 as well. Staff from one facility requested that SNF/Assisted Living be added as a response to N20. Staff from 3 of the 9 sites reported some difficulty classifying patients into only one of the responses offered for N21. One facility commented that patients and families rarely fall into one category or another. For example, many patients are home alone during the day or night but have primary caregiver assistance for the other part of the day. Additionally, some patients receive 24-hour assistance at a SNF or Assisted Living Facility, which is not an option. Facilities offered several suggestions for how to modify responses to N21. One facility suggested grouping answers by the number of hours a patient has received assistance rather than number of times per day or week, while another facility suggested adding a category for "several times during the night" or "several times during the day."

Staff from 4 of the 9 sites reported that there was no appropriate response for N21 if the response to N20 was "self." The IRF-PAI manual instructed staff to mark "unknown" for N21 when N20 was "self" and this suggestion did not seem logical to staff at several facilities. Several suggestions were offered for how to revise this: one facility suggested leaving N21 blank when N20 was 0 while another facility suggested adding a skip statement so if N20 was marked "self" item N20 would be skipped. One facility suggested adding a "none" category to N20 in order to indicate that the patient requires primary caregiver assistance and that no one is available, as opposed to answering "self" to indicate that the patient is able to take care of themselves

4.4.2 Changes to Question in the Medical Needs Section

Revisions were also made to the "Medical Needs" section of the IRF-PAI. During the first TEP, panel members agreed that the comatose item (item 25) was not necessary as it did not apply to IRF patients (i.e., a patient would not be admitted for rehabilitation if in a coma). This item was therefore removed from the revised version of the IRF-PAI. The dehydration item (item 28) was also deleted, as the TEP agreed this item was not necessary.

Delirium—TEP members suggested replacing the item 26 on delirium with a three-part question that evaluates if the patient is oriented to person/self, place, and time. Items N27A through N27C were added to the revised instrument to address this request (***Table 9***). About 98.3% of respondents were oriented to self at admission while 89.9% were oriented to place and 85.4% to time. There was only one missing response for this item. While "yes" responses were very high for all three questions, there seems to be some drop-off in "yes" responses, which captures some variation in orientation among patients. There were no suggested revisions for these three items (from staff at pilot test sites).

Table 9 (N27A-C)
Medical needs: Is patient oriented to self, place, and time?

	Percent		
Oriented	Self	Place	Time
0 – No	1.8	10.1	14.6
1 - Yes	98.3	89.9	85.4

Sample size = 514

Frequency missing = 1

Percentages may not add to 100 due to rounding error.

Swallowing—The swallowing item on the IRF-PAI was retained (N28B on the revised IRF-PAI), but an additional item was added for the pilot test (N28A). The Swallowing Functional Communication Measure (FCM) was adapted from American Speech Language and Hearing Association (ASHA) National Outcomes Measurement System (NOMS) and added to address TEP members' requests for a better item for assessing swallowing status (***Table 10***). This is a 7-level item used to determine a patient's swallowing ability using guidelines for dietary restrictions, cueing, and use of tube feeding and oral supplements. For the ASHA scale (item N28A) at admission, 75.2% of patients were coded as Level 7—the ability to eat was not limited by swallowing function; and 8.5% were coded as Level 6. The rest of the responses for this item (levels 5 through 1) accounted for slightly more than 15% of the total responses, with levels 4 and 5 receiving about 5.5% each. Overall, responses clustered into the top 3 levels.

Table 10 (N28A)
Medical needs: Swallowing measure

	Percent	
Swallowing measure	Admission	Discharge
7 – No limitations	75.2	81.4
6	8.5	7.1
5	5.6	4.7
4 – Swallowing is safe	5.2	4.2
3	2.7	1.0
2	0.2	0.6
1 – Not able to swallow	2.5	1.0

Sample size = 515 admission, 506 discharge

Frequency missing = 0 admission, 9 discharge

Percentages may not add to 100 due to rounding error.

The second swallowing item included in the pilot test (N28B) is the one currently used on the IRF-PAI **(*Table 11*)**. At admission, 82.1% of patients were coded as level 3—solids and liquids swallowed safely—while 14.4% were coded at Level 2 (modified food) and 3.5% at Level 1 (tube/parenteral feeding). Comparing responses to N27A (ASHA's swallowing item) and N27B (current swallowing item) indicates that ASHA's 7-level item could be collapsed into 3 categories similar to the ones on the current item with similar results. Approximately 83% of participants were classified into levels 6 or 7, similar to the 82% classified into Level 3 on the current item. In addition, collapsing levels 3, 4 and 5 into one category would create a combined proportion of 14%, similar to the 14.4% classified into level 2 of the old item.

Table 11 (28B)
Medical needs: Swallowing status

	Percent	
Swallowing status	Admission	Discharge
3 - Regular food	82.1	85.6
2 - Modified food consistency/supervision	14.4	12.3
1 - Tube/parenteral feeding	3.5	2.2

Sample size = 513 admission, 506 discharge

Frequency missing = 2 admission, 9 discharge

Percentages may not add to 100 due to rounding error.

For both the ASHA swallowing item (N28A) and the current item (N28B), results do not differ significantly for responses coded at discharge. A slightly higher proportion of patients were classified as Level 7 (or Level 3) and, as a result fewer patents were classified into the additional 6 levels. The completion rates for both ASHA's swallowing item and the current item were very high; only 9 responses were missing for each.

Comments on swallowing items received from staff during the pilot test were mixed. Many of the negative comments arose from difficulties encountered when coding the ASHA 7-level item. One facility mentioned that it was often difficult to obtain accurate cueing information during the initial swallowing evaluation at admission because it is necessary to watch the patients eat a meal. Staff at another facility had difficulty distinguishing between Level 3 and 4 and commented that scoring a patient on one of these levels rather than the other amounts to an educated guess.

Several positive comments were also forthcoming. Staff from two of the nine facilities felt that the ASHA scale was an improvement over the current item. One facility reported that it reflected the progress of patients in swallowing better than the current item, which reflects only the lowest level of swallowing for each patient. Another facility reported that the current item is not descriptive enough to accurately measure swallowing improvements that occur between admission and discharge.

There were also several comments about the way the ASHA swallowing item was presented on the IRF-PAI form. Staff from one site felt the definitions on the IRF-PAI form were not helpful in determining the actual scores and the manual was necessary when scoring, while staff from another site had the opposite reaction—definitions in the manual were too detailed and the descriptions on the IRF-PAI form were more useful than the training manual. Staff at several sites felt they would have benefited from more training on how to score the item. In general, nurses were less comfortable using this item than speech pathologists, although often the nurse was the primary assessor, unless a speech pathologist was assigned to the case.

4.4.3 Pre-Morbid Functional Status: Additions to the Function Modifiers/FIM Instrument

The "Goal" checkboxes on the FIM (item 39) were replaced with "3 Months Prior to Onset". Replacing this item addressed TEP members' request that a measure of premorbid physical functioning be added to the IRF-PAI (***Table 12***). The "Prior to Event" checkbox was added because panel members decided that any scale used to measure physical functioning should be comparable to scores obtained by the FIM so that data can be compared at admission and discharge.

Scores were collected through several methods. The directions in the manual asked the assessor to identify the patient's most typical dependent level during the 3 months prior to the onset of the condition being treated. The definitions for determining onset were the same as used in the IRF-PAI FIM's section. However, the data had to be collected through an interview (either asking the patient or family member during the stay). If they were unable to answer, the assessor was directed to review the patient's records for information or ask the psychiatrist. This is a very imprecise measure because the assessment is dependent largely on self-report. Comments suggested that staff could determine if the patient were independent, moderately dependent, or completely dependent prior to the acute event, but they found it difficult to apply as precise a scale as the 7-point FIM scale.

Second, staff at several facilities were worried that data reliability may vary by type of patient. For example, assessing functional levels three months prior to onset of impairment may be difficult for patients with head injuries and cognitive impairments, but they also commented that it is possible when family members are present. One facility recommended including an option for "cannot assess at this time" or "information not available" for 3 months prior FIMs questions.

Three of the nine IRFs commented specifically on the FIM cognition questions and felt that the scores may not be accurate when obtained by self-report from the patients. Staff at one facility were concerned that some patients were not honest with their self-report responses in this category. One facility suggested asking indirect questions to assess cognition, such as what types of tasks patients do at home (e.g., paying bills, finances, shopping).

Table 12 (N39A)
FIM instrument: 3 months prior to onset

	Percent								
FIM instrument	0	1	2	3	4	5	6	7	Missing
SELF-CARE									
a. Eating	0.2	0.2	0.4	0.2	1.0	5.3	11.1	81.7	2
b. Grooming	0.2	0.4	1.0	0.8	2.0	4.5	8.8	82.5	2
c. Bathing	0.2	0.6	2.7	1.6	6.1	6.1	16.6	66.2	3
d. Dressing - upper	0.2	0.6	1.8	0.4	2.3	5.5	9.0	80.3	3
e. Dressing - lower	0.2	0.8	2.9	1.4	4.5	4.7	13.7	71.9	3
f. Toileting	0.4	1.6	1.4	1.0	2.7	4.1	16.2	72.7	3
SPHINCTER CONTROL									
g. Bladder	0.4	3.3	0.8	0.8	2.2	2.9	9.6	80.0	4
h. Bowel	0.4	2.7	0.6	0.2	1.6	1.6	20.0	73.0	4
TRANSFERS									
i. Bed, chair, wheelchair	0.2	0.8	0.6	1.2	1.2	2.5	26.8	66.8	3
j. Toilet	0.2	1.2	0.4	1.4	1.6	1.8	30.7	62.9	3
k. Tub, shower	1.2	1.0	0.8	1.8	5.5	5.1	28.1	56.6	3
LOCOMOTION									
l. Walk, wheelchair	0.6	1.6	3.9	0.4	0.8	4.3	35.2	53.2	4
m. Stairs	7.3	6.1	7.7	0.4	1.2	3.4	33.5	40.4	8
COMMUNICATION									
n. Comprehension	0.2	0.0	0.0	0.2	2.2	4.1	32.8	60.6	5
o. Expression	0.2	0.0	0.0	0.4	1.2	3.7	13.7	80.8	5
SOCIAL COGNITION									
p. Social interaction	0.2	0.0	0.6	0.2	1.6	5.1	16.3	76.1	5
q. Problem solving	0.2	0.0	0.4	0.4	2.2	10.2	20.2	66.5	5
r. Memory	0.2	0.0	0.2	0.8	3.7	9.6	22.0	63.5	5

Sample size = 515

Percentages may not add to 100 due to rounding error.

4.5 Quality Indicators

Respiratory Status—The "Quality Indicators" section item 50 (the weak cough and difficulty clearing airways item) was removed, since TEP members did not feel it was necessary. However, the other two respiratory status items were retained (N48 and N49). These two items asked whether the patient had shortness of breath (SOB) with exertion or at rest at both admission and discharge. Less than 25% had SOB with exertion at admission and this number declined to only 15% by discharge, as one might expect after a rehabilitation therapy program (***Table 13***). The proportion having SOB at rest declined similarly during the stay, from 11.9% at admission to only 6.3% at discharge.

Table 13 (48-49)
Quality indicators: Respiratory status

	Percent	
Respiratory status	Admission	Discharge
Shortness of breath with exertion	24.5	15.0
Shortness of breath at rest	11.9	6.3

Sample size = 514 admission, 507 discharge

Frequency missing = 1 admission, 8 discharge

Percentages may not add to 100 due to rounding error.

Respiratory status may be difficult to assess for certain patients, such as those with high levels of brain injury. One facility commented that some staff members are unable to provide a truly accurate response to this item until head injuries resolve enough to allow patients to verbalize complaints about respiratory status. At this hospital, some patients have head injuries that are never able to resolve enough that patients can express respiratory problems.

Balance and Falls—The balance and falls items, 53 and 54, were both removed from the revised IRF-PAI. TEP members did not feel the falls item adequately assessed patient safety because patients are expected to fall during the rehabilitation process and if serious injuries resulted from a fall, the incident would be coded on the bills. During the first TEP, panel members also expressed that the balance item was not useful because it is too difficult to define and measure.

Pain—While the pain scale was retained as a result of TEP members' requests (renumbered to 50A on the revised IRF-PAI), an additional series of items were added to assess whether a patient's pain is being managed by the IRF and whether the pain is interfering with a patient's functioning (items N50B and N50C). Several TEP members suggested that a pain management item could be used as a process measure and should be field tested. Item 50A was not changed for the revised version of the IRF-PAI; however, staff from several facilities had comments on the scale's usefulness. Responses on the pain scale item were fairly well

distributed across the range of possible scores (***Table 14***). About 27.2% of patients reported no pain at admission, while 18.4% reported a moderate amount of pain (5 or 6 on the scale). Only 10% of patients reported severe pain (8 or 9). This reduced to no patients having a pain level of 10 by discharge.

Table 14 (50A)
Pain: Highest level of pain within assessment period

	Percent	
Pain rating	Admission	Discharge
0–No pain	27.2	34.0
1	1.4	6.3
2	3.1	6.5
3	3.3	6.3
4	6.8	5.5
5–Moderate pain	10.9	11.7
6	7.6	10.5
7	9.7	8.9
8	16.5	7.9
9	4.1	2.4
10–Worst possible pain	9.3	0.0

Sample size = 514 admission, 506 discharge

Frequency missing: 1 admission, 9 discharge

Percentages may not add to 100 due to rounding error.

Staff from 2 of the 9 pilot test sites reported encountering some difficulties in scoring this item. Since the item is self-reported, one facility had difficulty scoring the item when working with cognitively impaired or brain-injured patients. Staff was unable to assess the true level of pain as patients were unable to verbalize discomfort and staff was forced to look for nonverbal cues of pain. Some facilities had difficulty answering the pain scale even for patients without cognitive impairments. One facility indicated that the pain scale is not an accurate reflection of the patient's overall pain since the question asks for the patient's highest level of pain within the assessment period. Patients may experience a high level of pain during certain times of the day or certain activities, but when the activity is over or pain management has been done, they are pain-free for the rest of the day. The coordinator at this facility recommended rephrasing this question to ask for the general level of pain across the day. Most facilities reported they already used some method of recording pain level, and many were similar to the 7-point FACES tool (discussed in the literature review).

Item N50B assessed the method(s) employed by the IRF staff to manage a patient's pain during their stay and was to be collected at admission and discharge (***Table 15***). The majority of patients receiving pain treatment, received medications (75%), although a substantial number also may have received therapeutic massage or other modalities (17.1%). Only half the cases receiving psychological management at admission were still receiving it at discharge (3.7% and 1.8%, respectively.) About 13% of patients received some "other" form of pain management during the stay as noted at both admission and discharge.

Table 15 (N50B)
Pain: What methods for pain management are being used during the stay? (check all that apply)

	Percent	
Pain management method	Admission	Discharge
0 - Patient has no pain	28.5	30.9
1 - Medication	74.0	69.5
2 - Therapeutic massage	17.1	18.1
3 - Psychological management	3.7	1.8
4 - Other	13.0	13.6

Sample size = 515

Categories are not mutually exclusive.

There was some confusion among IRF pilot test participants over how to answer item N50B. This question refers to types of pain treatments during the stay but the information is recorded at admission and discharge. In addition, one IRF reported that staff from different specialties rated pain differently; where the PT felt a patient had no pain, the nursing staff felt the patient did have pain and administered medications. One facility recommended asking a two-part question: does the patient have pain (yes/no), and if yes, check all that apply. Two facilities recommended adding a question that assesses if a patient's pain limits their ability to function as a participant in therapy.

Item N50C assessed whether the patient's pain limited their ability to participate in the self-care process (***Table 16***). At admission, 29.6% of patients had pain great enough to interfere with the self-care process; this proportion decreased to 16.3% at discharge. There was only one comment reported by pilot test sites for item N50C. One facility did not like the responses available for the question and felt the yes/no responses were not descriptive enough to capture the full range of patients. They felt that patients may have pain that limits participation in the rehabilitation process all the time, part of the time, or never, leading to a 3-response answer.

Table 16 (N50C)
Pain: Does pain limit the patient's ability to participate in the self-care process?

	Percent	
Pain limiting?	Admission	Discharge
0 - No	70.5	83.7
1 - Yes	29.6	16.3

Sample size = 511 admission, 503 discharge

Frequency missing = 4 admission, 12 discharge

Percentages may not add to 100 due to rounding error.

Pressure Ulcers—The three pressure ulcer items currently on the IRF-PAI were evaluated. The TEP discussion raised questions about the appropriateness of these items for measuring quality in the IRF, since patients having ulcers are often admitted with them and this would, therefore, not be a fair measure of the quality of the IRF care. However, they also felt that a pressure ulcer should not worsen. The first item (51A) showed that the number of Medicare IRF admissions having pressure ulcers were few (84.6% had none at admission) and that it grew to 87.2% of discharges (***Table 17***). Of those with an ulcer, the highest current ulcer stage was most frequently partial loss of skin layers (48% at admission) which declined to only 45% at discharge (***Table 18***). The PUSH tool was retained as a measure because the wound care nurses pointed out that one could not see changes in wound severity without recording the length of the wound, the exudation amount, and the changes in the tissue type. Together these items are summed to create a total PUSH score. ***Table 19*** shows that the total PUSH scores for the Medicare IRF admissions decreased during the stays. The percent having a PUSH score of "0" increased from 68.9% at admission to 77.6% at discharge.

Table 17 (51A)
Pressure ulcers: Number of current pressure ulcers

	Percent	
Number of ulcers	Admission	Discharge
0	84.6	87.2
1	8.3	7.4
2	4.7	3.2
3	1.3	1.6
4	0.6	0.5
5	0.4	0.0
6	0.0	0.0
7	0.2	0.2

Sample size = 515

Percentages may not add to 100 due to rounding error.

Table 18 (51B)
Pressure ulcers: Highest current pressure ulcer stage

	Percent	
Pressure ulcer stage	Admission	Discharge
1 - Any area of persistent skin redness	33.3	31.7
2 - Partial loss of skin layers	48.0	45.0
3 - Deep craters in the skin	4.0	11.7
4 - Breaks in skin exposing muscle or bone	4.0	5.0
5 - Not stageable	10.7	6.7

Sample size = 75 admission, 60 discharge

Percentages may not add to 100 due to rounding error.

Table 19 (51F)
Pressure ulcers: Total PUSH score

	Percent	
Total PUSH score	Admission	Discharge
0	68.9	77.6
1-5	11.9	6.3
6-10	13.5	9.2
Greater than 10	5.7	6.9

Sample size = 515

Frequency missing (admission) = 322

Frequency missing (discharge) = 341

Percentages may not add to 100 due to rounding error.

Mood/Depression—RTI identified two items that were included in the revised IRF-PAI to assess mood or depression. These new items address TEP members' request that the IRF-PAI include an assessment of a patient's mood and/or depressive state as a case-mix variable rather than to form the basis of a quality indicator. Since there was no consensus from TEP members about which item was "best," both items were field tested. The first mood item (N52) was adapted from the RIC Functional Assessment Scale (RIC-FAS) and uses a 7-point rating scale intended for use by IRF staff to determine any disturbances of mood experienced by the patient. The second item (N53A through N53D) was adapted from the 1-item and 4-item mood assessment questions on the Geriatric Depression Scale (GDS). An additional item (N54) was added as a process measure to assess whether or not IRF staff referred patients with a higher score on mood item for a full evaluation. Several TEP members suggested that this item should be field tested because it could indicate whether patients are receiving necessary mental health services, which might be considered as a process quality indicator.

RIC-FAS Mood/Depression Item—At admission, over 85% of responses to the RIC-FAS mood item (***Table 20***) were clustered into the top three levels (5 through 7); 37.6% of patients reportedly had no evidence of depression, while 34.0% and 14.6% had a minimal or mild problem, respectively. Also, 2.9% of the patients were reported as having a moderate problem while slightly more than 2% were classified into the two more severe categories (levels 1 and 2). In 2% of the cases, the patient was not assessed for depression. These results did not change much when the patient was assessed at discharge; 44.5% of patients were reported to have no evidence of depression, up from 37.7% at admission. Scores for the rest of the scale shifted slightly in response to this increase. It is interesting that the proportion of individuals coded with a severe problem increase slightly, from 0.6% at admission to 1.0% at discharge. In addition, the proportion of patients not assessed decreased at discharge.

There were very few comments from pilot test sites about the RIC-FAS depression item. The comments that were received compared staff's experience in administering both depression items (the GDS and RIC-FAS). Two of the 9 facilities reported that the RIC-FAS scale was more descriptive than the GDS. One facility felt that the scale was too open to interpretation.

Table 20 (N52)
Mood and Depression: Lowest signs of depression within assessment period

	Percent	
RIC-FAS depression rating	Admission	Discharge
7 - No problem	37.7	44.5
6 - Minimal problem	34.0	31.4
5 - Mild problem	14.6	12.5
4 - Mild to moderate problem	6.6	6.7
3 - Moderate problem	2.9	2.0
2 - Moderate to severe problem	1.8	1.4
1 - Severe problem	0.6	1.0
0 - Not assessed	1.9	0.6

Sample size = 515 admission, 506 discharge

Frequency missing = 0 admission, 9 discharge

Percentages may not add to 100 due to rounding error.

Geriatric Depression Scale Item—The GDS items (N53A through N53D) are self-report items where staff must interview the patient to collect the data (***Table 21***). This item was included because it is frequently used with the geriatric population and was suggested as a possible appropriate item. Staff was instructed to ask the patients to choose the best answer for how they felt in the past week. If the answer to the first question (N53A) was "no," then staff were to skip the rest of the questions and enter "0" for N53E. If the patient responded "yes" to the first question, staff was to continue and ask N53B through N53D and enter a composite score in N53E. The frequencies for this set of items indicate that facilities had some difficulty

completing the items. Many staff commented in the trainings that they were uncomfortable asking a patient if they felt like their "life was empty," and as a result, many skipped the first item and went on to ask items b-d.

Table 21 (N53-N54)
Mood and depression: Geriatric depression scale

	Percent "Yes"		Missing	
GDS depression scale	Admission	Discharge	Admission	Discharge
53.a. Life empty?	8.3	4.5	41	48
b. Satisfied with life?	53.9	45.0	476	495
c. Afraid something is going to happen?	28.2	35.0	476	495
d. Happy most of the time?	85.8	84.1	346	389
f. Can patient answer questions?	94.6	94.6	11	34
54. Patient referred to mental health?	16.3	14.6	12	34

Sample size = 515

Categories are not mutually exclusive.

At admission, 8.3% of patients (41 individuals) answered "yes" to the first question. Given this, items N53B through N53D should also have 41 responses. However, only 39 patients had responses coded for N53B and N53C. In addition, item N53D has a response coded for 169 individuals, indicating that staff that received a "no" on the first question may have skipped N53B and N53C, and answered N53D even though they were instructed to skip all three items. Given the apparent problems encountered in answering these items, the composite scores reported in N53E are difficult to interpret. Since only 41 individuals answered item N53A "yes," there should be a composite score greater than 0 for only 41 patients (***Table 22***). However, 52 patients had a composite score of 1 or more. This discrepancy may indicate that staff had difficulty totaling the composite score as well as answering the correct sequence of questions. Completion for this item may also be a problem; there were 21 missing values for the first question (N53A), which is necessary to complete the rest of the scale.

Table 22 (N53E)
Mood and Depression: Total depression score

	Percent	
GDS total mood and depression score	Admission	Discharge
0	89.9	94.0
1	4.7	3.5
2	2.7	0.8
3	1.6	0.8
4	1.2	1.0

Sample size = 52 admission, 52 discharge

Percentages may not add to 100 due to rounding error.

At discharge, 4.5% of patients (21 individuals) answered "yes" to the first question (N53A). Frequencies indicate continuing confusion about how to answer the following three questions (N53B through N53C) as only 9 patients answered N53B; 7 answered N53C; and 126 N53D. All three questions should have 21 responses coded. The number of missing responses also increased to 48 at discharge.

An additional item (N53F) was added to assess whether or not a patient could answer the 4 self-reported GDS items included on the revised instrument. At admission 5.4% of patients (27 individuals) were reportedly unable to answer the GDS items, which may account for the lower completion rate and some of the missing responses reported above. However, there were 11 missing responses for this item alone.

Comments received about the GDS items from pilot study participants were generally negative; 5 of the 9 sites reported that the GDS item was difficult to score for patients with cognitive impairments due to brain injury or stroke. Sites commented that patients were unable to respond, were confused by the questions, or had other mental health problems related to brain injury rather than true depression. Staff at several sites reported that some cognitively impaired patients were able to answer the GDS questions, but questioned the reliability of the answers. Patients with communication problems caused by brain injury or stroke were also unable to answer the GDS questions.

Staff at 4 of the 9 sites reported being uncomfortable asking question N53A ("Do you feel that your life is empty?"), which could account for the large number of missing responses for this item. One site reported that alert and oriented patients wondered why they were being asked this question. One neuropsychologist commented that this question has little practical meaning for the patient and that most patients answer "no" if they do not understand the question.

One site reported encountering difficulties obtaining information for the GDS since multiple therapists and nurses treat each patient and the patient does not present the same symptoms of mood and depression to each therapist. Staff at this facility suggested that mood and depression questions should only be scored at discharge.

Engagement—An engagement item was added to the revised IRF-PAI in order to measure a patient's involvement in treatment (N55). TEP members felt that a patient's level of engagement could be an important risk adjustor, influencing how effective potential rehabilitation interventions might be. This item was adapted from the Rehabilitation Institute of Chicago (RIC) Functional Assessment Scale (RIC-FAS) and uses a seven-point rating scale assessed by staff working with the patient in the IRF. The word "classes" was replaced with "therapy sessions" in the coding for levels 2, 4, and 5. Adding this item addressed the TEP members' request that an engagement item be included on the IRF-PAI in order to account for the wide range of factors that influence a patient's level of participation in rehab.

At admission, 42.1% of patients were coded as having no problem with engagement/participating in the rehab process (***Table 23***). About 31.4% reportedly had minimal problems and 11.9% had a mild engagement problem. The top three levels account for 85.5% of all patients in our sample. About 10% of patients were reported to have moderate problems (levels 3 and 4) while 3.5% of patients had more severe problems (levels 1 and 2). Only 4

patients were not assessed and there were only 2 missing responses in the sample. At discharge, scores shifted slightly, with 50.2% of patients reportedly having no problem with engagement, compared to 42.1% at admission. Proportions in levels 5 and 6 decreased slightly, to 26.5% and 10.3%, respectively, at discharge. Scores on the remaining 4 levels were similar to those at admission. Only 2 patients were not assessed, but the frequency of missing responses increased to 9.

Staff at pilot test sites encountered some difficulties in administering the engagement item, though there were no consistent concerns reported about the item across sites. Similar to the mood and depression items, staff at one facility felt that the rating scale for engagement was not compatible with brain injury and stroke survivors with cognitive difficulties or communication problems. Patients with a brain injury often do not understand the value of therapy and its contribution to the recovery process. These patients would be given the lowest score on the engagement scale due to impaired cognition rather than actual problems with engagement. Staff at another facility felt the engagement item was not applicable to patients who are unaware they are receiving therapy, and suggested the use of multiple engagement items for different types of therapy, as patients may be less engaged in particular therapies that they dislike. Staff at one facility had difficulty assessing engagement due to multiple therapists and nurses seeing each patient. In these situations, different staff may have different perceptions of a patient's engagement.

Table 23 (N55)
Engagement: Patient's cognitive and emotional resources to actively participate in program

	Percent	
Engagement rating	Admission	Discharge
7 - No problem	42.1	50.2
6 - Minimal problem	31.4	26.5
5 - Mild problem	11.9	10.3
4 - Mild to moderate problem	5.1	5.3
3 - Moderate problem	5.3	3.6
2 - Moderate to severe problem	2.5	2.8
1 - Severe problem	1.0	1.0
0 - Not assessed	0.8	0.4

Sample size = 513 admission, 506 discharge

Frequency missing = 2 admission, 9 discharge

Percentages may not add to 100 due to rounding error.

Staff at all facilities was pleased to see an engagement item included in the IRF-PAI. Across all disciplines, this item was considered one of the most important factors for measuring patients' expected functional changes and probability of being discharged home. The comment was also made repeatedly that keeping patients engaged, or dealing with engagement issues, is one of the factors that is never recorded, yet is an important part of the rehabilitation process.

Staff was unanimously pleased to see this item proposed for the new dataset. A few commented on the overlapping among depression, mood, and engagement. It was felt that depression and mood are underlying factors, among others, that may affect engagement levels and that engagement was the more important measure in evaluating outcomes.

SECTION 5
DEVELOPING QUALITY INDICATORS FOR INPATIENT REHABILITATION FACILITIES RISK ADJUSTMENT MODELS

5.1 Analytic Objectives

One of the primary reasons for CMS' desire to identify and test potential new items for a revised IRF-PAI instrument (discussed in Section 4) was the goal of developing potential quality indicators for inpatient rehabilitation facilities. While the primary focus of the IRF-PAI instrument and the data collected is to support the mandated IRF PPS, data collected from this instrument could also support a parallel quality monitoring system. In addition to developing the basic quality outcome measures, it is critical to develop corresponding facility risk adjustment models that would be used in the proper interpretation of the outcome measures, and would allow for appropriate comparisons across facilities. Therefore, RTI conducted some preliminary work to identify potential IRF quality indicators as well as risk adjustment models.

The Technical Expert Panel (TEP) selected two outcome criteria for measuring the impact of inpatient rehabilitation services: change in a patient's functional status and whether the patient was discharged to the community (defined as home, board and care, assisted living, and transitional living). These two outcomes were identified because they are often considered the primary goals of any inpatient rehabilitation treatment program. The objective of the analyses presented in this section was to identify risk adjustment factors that would control for patient differences across facilities while examining these outcomes. It is imperative to note that these analyses are only exploratory in nature. Given the timeline and concerns of CMS, the data on which these analyses were conducted represent: 1) a small sample of hospital admissions; while the types of cases included are diverse, the hospitals are not a nationally representative sample; 2) short time frame for data collection and analysis to meet CMS' internal time constraints, and 3) interpretation of the results must take into consideration the many questions and issues the IRFs raised regarding some of the items being tested on the revised form, including questions about how to complete an item and what information a particular variable captured. While the sample available to us for these analyses was small, the results are indicative of the types of risk adjustment factors that are associated with different rehabilitation outcomes warranting further research on a larger scale.

This section presents information on which factors are associated with changes in functional improvement (both motor and cognitive) or with the probability of discharge to the community. In some cases, multiple items have been pilot tested to identify the best measure of a concept as it relates to these outcomes. First, this section presents an overview of the sample populations analyzed and the variables included in the models. Second, basic models using only existing IRF-PAI variables are presented to compare the three samples. Third, pilot test items are included in stepwise Ordinary Least Squares (OLS) regression models and logistic models that are used to identify, respectively, which factors contribute to the various outcomes and the relative contribution of these new pilot-tested measures in predicting FIM changes or discharge to community.

5.2 Data Sources

Three datasets were utilized for the analyses. First, we compared elements common to both the national IRF-PAI dataset and the pilot dataset to examine how closely the pilot data reflects the national population. Second, we present data from the nine hospitals involved in the pilot test to get a sense of how much of the differences between the national and pilot data are due to differences between the nine participating hospitals and the national population. The remaining analyses focus on the pilot data to identify important potential factors for risk adjusting quality outcome measures.

The first dataset, "National Sample" (N = 494,606 admissions), refers to a subset of the national data collected from the original IRF-PAI form utilizing the following inclusion criteria for discharge patients:

- The Medicare Program was the primary or secondary payment source.
- The rehabilitation admission took place after July 1, 2002 and the discharge occurred no later than July 30, 2003.
- Treatment was received at an inpatient rehabilitation facility with a minimum of 30 admissions during the reference time period.

The second dataset analyzed is the "Nine Hospitals Sample" (N = 1,954 admissions). This dataset consists of a subset of the National Sample. Additional inclusion criteria for this sample are:

- Admissions from the nine inpatient rehabilitation facilities implementing the pilot study of the revised IRF-PAI form.
- Admission dates from May 1, 2003, and discharges no later July 30, 2003, representing the latest three-month period of admissions in the National Sample.

The third dataset, "Pilot Sample" (N = 515 admissions), refers to the data collected from the same nine inpatient rehabilitation facilities implementing the pilot test of the revised IRF-PAI form (i.e., the same facilities as in the second dataset) during the time period of October 2004 to January 2005. The data collected during the pilot test were merged with the IRF-PAI data collected from the original form and submitted by the hospitals to CMS for these admissions during the same timeframe (October 2004 to January 2005).

The data from the National Sample provides the most reliable estimates, but overall explanatory power is less than the Pilot dataset since the latter has richer variables than the National Sample. The Nine Hospital Sample allows us to understand selection bias relative to the nine facilities in comparison to the National Sample (i.e., generalizeability of the Pilot sample).

5.3 Data Steps

The basic steps in the analyses are as follows:

1. Perform Ordinary Least Square (OLS) regressions to examine change in patients functional status and logistic regressions to analyze the probability of discharge to community. These analyses will test the existing IRF-PAI form variables of interest for the National, Nine Hospitals, and Pilot samples.

2. Add new and/or revised variables from the pilot tested IRF-PAI form for the Pilot Sample to the basic models.

3. Analyze whether these new/revised variables add to the explanatory power of the models in order to recommend to CMS which pilot-tested items warrant further research and possible inclusion on the IRF-PAI form.

5.4 Outcome Variables

For the OLS regressions examining the functional status outcome, changes in a patient's FIM scores were examined. A number of new variables based on various combinations of FIM items were created by summing differences in FIM scores between discharge and admission. Higher, positive change scores indicate more gains in functional status. These FIM variables are based on the literature and suggestions of Dr. Margaret Stineman, a member of the TEP and project consultant.

"FIM Motor Change Score" was created by summing and combining all the differences in FIM scores between discharge and admission for the Self-Care, Sphincter Control, Transfers, and Locomotion FIM items.

"FIM Cognitive Change Score" was created by summing all the differences in FIM scores between discharge and admission for the Communication and Social Cognition FIM items.

"FIM Self-Care Change Score" was created by summing all the differences in FIM scores between discharge and admission for the Self-Care FIM items. This variable addresses the ability of the patient to undertake their activities of daily living (ADLs). This outcome item primarily focuses on the patient's upper body strength.

"FIM Sphincter Control Change Score" was created by summing all the differences in FIM scores between discharge and admission for the Sphincter Control FIM items.

"FIM Mobility Change Score" was created by summing all the differences in FIM scores between discharge and admission for the Transfer and Mobility FIM items.

In addition, the logistic regression models examined factors that predicted "Discharge to Community." This outcome was measured as patients who were discharged to a community setting versus to other settings. "Discharge to Community" includes the categories of discharge to "Home," "Board and Care," "Transitional Living" and "Assisted Living." "Other Setting"

includes the categories of “Intermediate Care/Long-Term Care,” “SNF,” “Acute Unit of Another Facility,” “Acute Unit of Own Facility,” “Chronic Hospital,” “Rehabilitation Facility,” “Other,” “Alternate Level of Care Unit,” and “Subacute Setting.”

Descriptive statistics comparing the FIM change scores outcome variables among the three datasets demonstrate relatively comparable FIM change scores (***Table 24***). All five FIM change scores were similar among the samples for mean and median values as well as standard deviation. Due to the effect of sample size, the Pilot Sample demonstrated the greatest standard error and the National Sample the least. The National Sample exhibited the greatest variability in the range of scores, as would be expected due to sample size.

Descriptive statistics comparing the three sample populations for the outcome variable “Discharge to Community” are shown in ***Table 25***. The samples were comparable, although slightly more patients were discharged to a community setting in the Pilot sample; 81% compared to the National and Nine Hospitals samples, with 77% and 78% of patients discharged to community settings respectively.

5.5 Independent Variables

The independent variables entered in the models to predict the outcomes of interest are Gender, Marital Status, Pre-Hospital Living Setting, Age, Charlson Index, various combinations of FIM admission scores, and Impairment Group. These variables have been important in past risk adjustment models for the Medicare population. Marital Status was collapsed into a dichotomous variable “Married” or “Other”; Pre-Hospital Living Setting was also collapsed into two categories “Community” (home, board and care, transitional living, assisted living residence) or “Other Setting.” Age was entered in the models as a continuous variable but is included in ***Table 26*** as a categorical variable.

The Pilot Sample contains more male patients than the Nine Hospitals and National samples, with 40%, 38% and 37%, respectively (Table 26). A higher percentage of patients in the Pilot Sample are married, 51% compared to 46% for the Nine Hospitals Sample and 44% for the National Sample.

The percent of individuals living in a community setting prior to their admission to a rehabilitation hospital was very similar among the Pilot, Nine Hospitals, and National samples: 97%, 97%, and 96%, respectively. Patients in the Pilot Sample tended to be younger; 22% of patients were 64 years of age or less compared to only 12% and 15% of the Nine Hospitals and National samples. Continuing this trend, only 7% of patients in the Pilot Sample were age 85 or older compared to 10% and 13% in the Nine Hospitals and National samples respectively.

The Charlson Index was utilized as a severity of illness index (http.www.umanitoba.ca/mchp/concept/dict/charlson.index.html) (see ***Table 27***). The Charlson Index contains 19 categories of comorbid conditions defined utilizing ICD-9-CM codes. Each category is associated with a weight based on the adjusted risk of one-year mortality. Thus, a higher Charlson Index score is associated with more severe comorbidity.

Table 24
Sample comparisons: Outcome variables, FIM change scores

Variable	Sample	Value					
		Mean	Median	Minimum Value	Maximum Value	Standard Deviation	Standard Error
FIM Motor Change							
	National	20.98	21	-53	78	11.84	0.02
	Nine Hospitals	21.97	22	-42	68	12.24	0.28
	Pilot	20.50	20	-17	59	11.06	0.49
FIM Cognitive Change							
	National	2.07	1	-30	30	3.95	0.01
	Nine Hospitals	2.28	1	-22	23	3.97	0.09
	Pilot	3.09	2	-15	20	4.43	0.20
FIM Self Care Change							
	National	8.34	8	-30	42	5.83	0.01
	Nine Hospitals	8.54	9	-23	31	5.82	0.13
	Pilot	8.16	8	-13	23	5.49	0.24
FIM Sphincter Control Change							
	National	2.34	2	-12	12	3.20	0.00
	Nine Hospitals	2.71	2	-11	12	3.27	0.07
	Pilot	2.17	1	-9	12	3.31	0.15
FIM Mobility Change							
	National	10.30	10	-22	34	5.57	0.01
	Nine Hospitals	10.71	11	-15	27	5.94	0.13
	Pilot	10.16	10	-4	29	5.35	0.24

Table 25
Sample comparisons: Outcome variable, discharge to community setting, frequency and percent

	National		Nine Hospitals		Pilot	
	Frequency	Percent	Frequency	Percent	Frequency	Percent
Discharge to Community	384,870	77.81	1,498	76.50	415	81.44
Discharge to Other Setting	109,736	22.19	456	23.28	100	18.56
Totals	494,606	100	1,954	100	15	100

Table 26
Sample comparisons: Selected independent variables, frequency and percent

	National		Nine Hospitals		Pilot	
	Frequency	Percent	Frequency	Percent	Frequency	Percent
Gender						
Male	182,220	36.84	743	38.02	208	40.39
Female	312,386	63.16	1,211	61.98	307	59.61
Marital Status						
Married	218,736	44.22	897	45.91	265	51.46
Other	275,870	55.78	1,057	54.09	250	48.54
Pre-Hospital Living Residence						
Community	476,469	96.33	1,899	97.19	499	96.89
Other Setting	18,137	3.67	55	2.81	16	3.10
Age Group						
0-64	60,234	12.18	286	14.64	111	21.55
65-69	71,851	14.53	307	15.71	76	14.76
70-74	96,902	19.59	359	18.37	97	18.83
75-79	111,265	22.50	478	24.46	113	21.94
80-84	89,998	18.20	331	16.94	80	15.53
85+	64,356	13.01	193	9.88	38	7.38

Table 27
Sample comparisons: Charlson Index, FIM admission scores

Variable	Sample	Value				Standard Deviation	Standard Error
		Mean	Median	Minimum Value	Maximum Value		
Severity Of Illness							
Charlson Index							
	National	1.21	1	0	14	1.42	0.00
	Nine Hospitals	1.16	1	0	10	1.44	0.03
	Pilot	1.18	1	0	9	1.46	0.06
FIM Score							
FIM Motor Adm							
	National	41.03	42	2	91	12.94	0.02
	Nine Hospitals	40.50	42	2	79	13.00	0.30
	Pilot	39.80	41	8	75	13.17	0.58
FIM Cognitive Adm							
	National	26.69	29	5	35	7.56	0.01
	Nine Hospitals	25.19	27	5	35	7.22	0.16
	Pilot	25.23	27	5	35	7.52	0.33
FIM Self Care Adm							
	National	22.43	23	0	42	6.65	0.01
	Nine Hospitals	22.38	23.	0	40	6.56	0.15
	Pilot	21.83	23	6	35	6.54	0.29
FIM Sphincter Control Adm							
	National	30.35	32	2	56	9.20	0.01
	Nine Hospitals	29.63	31	2	53	9.05	0.20
	Pilot	29.50	31	8	48	9.32	0.41
FIM Mobility Adm							
	National	10.67	10	0	35	4.91	0.01
	Nine Hospitals	10,87	10	0	33	5.11	0.16
	Pilot	10.30	10	0	29	4.93	0.22

The descriptive statistics for Charlson Index scores, among all the population samples, are comparable in regard to mean, median, standard deviation, and standard error. The Pilot Sample demonstrated the lowest maximum value, 9, compared to 10 and 14 for the Nine Hospital and National samples, respectively, suggesting that patients in the Pilot Sample may be slightly healthier than those the other two samples. A minimum value of 0 was consistent for all three samples.

FIM admission scores were included in the models as predictive of the final change in the FIM scores at discharge. Individual FIM item scores range from 1 (total dependence) to 7 (total independence). As stated earlier in this section in the discussion of FIM outcome variables, the FIM admission variables are various combinations of the individual FIM items. The admission mean and median values as well as standard deviation for the summed FIM Motor, FIM Cognitive, FIM Self-Care, FIM Sphincter Control, and FIM Mobility scores are very comparable across samples and all scores differed by two points or less. Standard errors were greatest for the Pilot Sample and smallest for the National Sample due to the size of the sample population. Generally, there were lower minimum and higher maximum FIM change scores in the National and Nine Hospitals samples compared to the Pilot Sample. The exceptions were the FIM Mobility Admission score for which the minimum value was zero for all three samples and the FIM Cognitive Admission Score, for which the minimum and maximum values were equivalent across the samples.

A new variable was constructed to create categories of impairments groups based on the work of Stineman et al. and TEP suggestions. Dr. Stineman's work on Functional Independence Staging (Stineman, Ross, Fielder, et al., 2003) collapsed Rehabilitation Impairment Categories (RICs) into six impairment groupings. Functional Independence Staging (FIS) "frames impairments in body function and structure by combining closely related RICs…..that more accurately reflect the International Classification of Functioning Disability and Health (ICF) definition of impairment" (Stineman, Ross, Fielder, et al., 2003). The impairments reflect the primary reason for the patient's admission to a rehabilitation facility. The six impairment groups utilized in Dr. Stineman's Functional Independence Staging research are: traumatic and nontraumatic damage to the central nervous system (CNS), traumatic and nontraumatic spinal cord dysfunction (SCD), other types of neurological involvement (Neurologic), impairment to the musculoskeletal system (Musculoskeletal), medical conditions that impair endurance or energy production (Endurance) and Other, a category that includes medical conditions involving multiple organs or that are not elsewhere classified. Additionally, as recommended by the TEP, Stroke and Replacement of Lower Extremity Joint were distinct categories given the large volume of admissions in those diagnostic categories. A set of dummy variables were created for the impairment groups. ***Table 28*** provides a crosswalk of RICs to Impairment Groups.

The frequency distribution of Impairment Groups among the sample populations can be found in ***Table 29***.

The breakdown of admissions by impairment grouping demonstrated that the Pilot Sample was slightly more similar, in frequency distribution, to the National Sample than to the Nine Hospitals Sample. The Pilot Sample was within three and one-half percentage points of the National Sample for all Impairment Group categories, and within three percentage points of the

Table 28
Crosswalk of RIC to Impairment Groups

RIC Name	Impairment Group	Description
Stroke	CNS	Central Nervous System damage
Traumatic Brain Injury	CNS	
Non-traumatic Brain Injury	CNS	
Traumatic Spinal Cord	SCD	Spinal Cord dysfunction
Non-traumatic Spinal cord	SCD	
Neurological	Neurologic	Neurological impairments
Hip Fracture	Musculoskeletal	Musculoskeletal impairments
Replacement of Lower Extremity Joint	Musculoskeletal	
Other Orthopedic	Musculoskeletal	
Amputation, Lower Extremity	Musculoskeletal	
Amputation, Other	Musculoskeletal	
Osteoarthritis	Musculoskeletal	
Rheumatoid, Other Arthritis	Musculoskeletal	
Cardiac	Endurance	Impairments affecting endurance or energy
Pulmonary	Endurance	
Pain Syndrome	Endurance	
MMT, No Brain or Spinal Cord Injury	Other	Other - medical conditions not elsewhere classified or affecting multiple organs
MMT, With Brain or Spinal Cord Injury	Other	
Guillain-Barre	Neurologic	
Miscellaneous	Other	
Burns	Other	

MMT = Major Multiple Trauma

Table 29
Sample comparisons: Impairment group, frequency and percent

	National		Nine Hospitals		Pilot	
	Frequency	Percent	Frequency	Percent	Frequency	Percent
IMPAIRMENT GROUP						
Central Nervous System (except stroke)	17,006	3.44	76	3.89	28	5.44
Endurance	48,593	9.82	302	15.46	41	7.96
Musculoskeletal (except LE joint replacement)	115,732	23.40	334	17.09	111	21.55
Neurologic	22,657	4.58	56	2.87	28	5.44
Other	68,540	13.86	279	14.28	90	17.48
Replacement of Lower Extremity Joint	119,078	24.08	545	27.89	114	22.14
Spinal Cord Dysfunction	20,055	4.05	107	5.48	23	4.47
Stroke	82,945	16.77	255	13.05	80	15.53

Nine Hospitals Sample for all Impairment Groups with the exception of Endurance (7.5% difference), Musculoskeletal (4.5% difference), and Replacement of Lower Extremity Joint (5% difference).

5.6 Methods Summary

Initially, “basic” models for each outcome were established fixing existing IRF-PAI variables in multivariable regression models. Next, beginning with those basic models, we examine the potential of the proposed quality indicators to explain each outcome through forward stepwise selection starting with the basic models.

5.6.1 Basic Risk Adjustment Models for FIM Change Scores

The next step in the analyses consisted of building basic models using the variables contained in the current version of the IRF-PAI form for all three datasets. The basic models include demographic variables (gender, marital status, pre-hospital living setting and age), impairment group, severity of illness, and FIM admission scores using a fixed effects mode. The variable "Married" was used as an independent variable instead of "Pre-Hospital Living With" due to the fact that the latter variable was only completed if the patient was living at home. The majority of variables in the Quality Indicators section of the IRF-PAI form could not be included in the basic model analyses due to the under-reporting of these items as voluntary, non-test items. However, many of these variables are tested in the next section using only pilot data.

Table 30 provides a summary of the five basic ordinary least squares (OLS) regression models for all three samples displaying the adjusted R-squared for each model and identifying whether the independent variables were significant, and if so, the direction of the correlation coefficient.

National Sample. For all five models, with two exceptions, all independent variables were significant at the P = <0.001 level. The exceptions were for the variable "Married" in the models examining the change in FIM Cognitive and FIM Sphincter Control scores where the variable was significant at the P = 0.05 level. The models with the greatest explanatory power were the Cognitive FIM Change and Sphincter Control FIM Change with adjusted R-squared values of 0.24 and 0.33, respectively. The model with the least explanatory power was the FIM Self-Care Change with an adjusted R-squared value of 0.12.

Next, beta coefficients were examined to determine their relative contribution to the models. Being female was positively associated with greater FIM change score (i.e., greater gains in functional status) for all the models with the exception of FIM mobility change. Whether a patient was married was negatively associated with a change in FIM score for all but the model examining Cognitive FIM Change. In this subscale being married was positively associated with an increase in the change score. Being married may be associated with a patient's need to be somewhat less independent. Living in the community prior to admission to rehabilitation facility (Pre-Hospital Living Setting) was associated with higher FIM change score for all five models. Age was negatively associated with the FIM change scores for all the models.

The Impairment Group variables were significant in all the models. The "Replacement of Lower Extremity Joint" Impairment Group was utilized as a reference group for the models. Relative to this group, all other impairment groups were negatively associated with change in FIM scores. This impairment group was found to be the "healthiest" in terms of Charlson Index and admission FIM scores, and were a younger population. A higher Charlson Index (greater morbidity) was negatively associated with a FIM change score. Intuitively, it makes sense that patients who are older and with more comorbidities would experience less gains in the various FIM change scores.

Table 30
Sample comparisons: Preliminary basic risk adjustment models, OLS regression, FIM change score

Dependent Variables	FIM MOTOR CHANGE			FIM COGNITIVE CHANGE			FIM SELF CARE CHANGE			FIM SPHINCTER CONTROL CHANGE			FIM MOBILITY CHANGE		
Independent Variables	National	Nine Hospitals	Pilot	National	Nine Hospitals	Pilot	National	Nine Hospitals	Pilot	National	Nine Hospitals	Pilot	National	Nine Hospitals	Pilot
Demographic Variables															
GENDER (Female)	+	NS	NS	+	NS	NS	+	NS	NS	+	+	+	--	--	NS
MARRIED (Yes)	--	NS	NS	+	NS	NS	--	NS	NS	--	NS	NS	--	NS	NS
PRE-HOSPITAL LIVING SETTING (Community)	+	NS	NS	+	NS	NS	+	NS	NS	+	NS	NS	+	NS	NS
AGE	--	NS	NS	--	--	NS	--	NS	NS	--	NS	NS	--	NS	NS
Impairment Group															
CENTRAL NERVOUS SYSTEM	--	NS	NS	--	--	--	--	NS	NS	--	NS	NS	--	NS	NS
SPINAL CORD DYSFUNCTION	--	--	--	--	NS	NS	--	--	--	--	--	--	--	--	NS
NEUROLOGIC	--	--	--	--	NS	NS	--	--	--	--	NS	--	--	NS	--
MUSCULOSKELETAL	--	--	NS	--	--	NS	--	--	NS	--	NS	NS	--	--	--
STROKE	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--
ENDURANCE	--	--	--	--	--	--	--	--	--	--	NS	--	--	--	NS
OTHER	--	--	NS	--	--	--	--	--	NS	--	--	NS	--	--	NS
Severity of Illness															
CHARLSON INDEX	--	--	--	--	--	NS	--	--	NS	--	--	NS	--	--	--
FIM Items															
FIM COGNIT ADM	+	+	NS	--	--	--	+	+	+	+	+	+	+	+	NS
FIM SELF-CARE ADM	--	NS	NS	+	+	NS	--	--	--	+	+	+	+	+	+
FIM SPHINCTER CONTROL ADM	--	--	NS	+	NS	NS	+	+	+	--	--	--	+	+	+
FIM MOBILITY ADM	--	--	--	+	+	NS	+	+	+	+	+	+	--	--	--
R-squared	0.08	0.13	0.09	0.24	0.21	0.34	0.12	0.15	0.19	0.33	0.38	0.33	0.15	0.22	0.20
Adjusted R-squared	0.08	0.12	0.07	0.24	0.21	0.32	0.12	0.14	0.16	0.33	0.37	0.31	0.15	0.21	0.17
N	494,606	1,954	501	494,606	1,954	501	494,606	1,954	501	494,606	1,954	501	494,606	1,954	501

NOTES: ( ) Denotes reference group in the analysis
The impairment group "Replacement of Lower Extremity Joint" is the reference group in the analysis.
Significance Level = 0.10 + = Significant positive effect -- = Significant negative effect NS = Not significant

Finally, FIM admission scores were significant but varied in direction depending on the outcome measured. The FIM admission scores corresponding to the dependent variable were negatively associated with a change score. That is, patients admitted with higher FIM admission scores experienced less change in their score because they were functioning at a higher level at the beginning of their rehabilitation stay. The addition of more difficult functions such as instrumental activities daily living (IADLs) could provide the means to capture the achievements of less functionally severe patients. The one exception was for the model examining the change in the FIM Motor score. The independent variables taking into account FIM Self-Care, Sphincter Control, and Mobility scores on admission for these FIM categories were associated with lower FIM Motor Change scores. One plausible explanation is that all three FIM categories comprise the FIM Motor score.

The models were then run using the Pilot Sample and the adjusted R-squared values for three of the five models (FIM Motor Change, FIM Self-Care Change, and FIM Sphincter Control Change) were comparable to those for the National Sample. The adjusted R-squared values for the FIM Cognitive Change model increased by 8 points from the National Sample results and the FIM Mobility Change model R-squared value increased by 5 points. Several of the independent variables became not significant at the P=0.1 level of significance. The independent variables that remained significant for some of the models were the Impairment Group variables and the Charlson Index. Their direction remained the same as for the results of the national models. Many of the FIM admission scores also dropped out the model as significant with two exceptions. The FIM admission items for the FIM Self-Care Change and FIM Sphincter Control Change models all remained significant and with the same direction. For all the models using the Pilot Sample, although some of the FIM admission variables became not significant, those remaining as significant did not change the direction of the beta coefficient.

The models were again run using the Nine Hospital Sample in order to analyze the similarities or differences compared to the Pilot Sample given the greater number of admissions in each hospital. The adjusted R-squared value (0.12) was greater than the National Sample (0.08) and Pilot Sample (0.07) for the FIM Motor Change model. For the Total Cognitive Change model, the adjusted R-squared value (0.21) for the Nine Hospital Sample was more similar to the National Sample (0.24) than the Pilot Sample (0.34). The adjusted R-squared value (0.14) for the FIM Self-Care model was similar to both the Pilot Sample (0.16) and National Sample (0.12). The adjusted R-squared value for the Total Sphincter Change model was also comparable for all three samples: Nine Hospital (0.37), National (0.33) and Pilot (0.31). Finally, the adjusted R-squared value for the Nine Hospital Sample (0.21) was more similar to the Pilot Sample (0.17) than the National Sample (0.15) for the FIM Mobility Change model.

Similar to the results of the regression runs using the Pilot Sample, several of the demographic variables fell out of the models and were not significant at the P=0.1 level. These results may be due to sample size differences between the National Sample and the other two datasets. The remaining significant independent variables were Gender (for the models examining FIM Sphincter Control and Mobility Change), Impairment Group, and Charlson Index for various models. Age remained significant only for the FIM Cognitive Change model. More FIM admission scores remained in the models than for the Pilot Sample. For the FIM Motor Change and FIM Cognitive Change models, there was a trend for some of the FIM admission scores to fall out of the models and become not significant, those remaining as

significant did not change direction of their beta coefficients. There was no change in the significance of the FIM admission scores for the FIM Self-Care, Sphincter Control, and Mobility Change models.

5.7 Pilot Test Models

The next step was to introduce the new IRF-PAI test items into the basic models to determine which variables contributed explanatory power in predicting the outcomes over and above the original IRF-PAI variables. The outcomes included the five FIM Change Score models and the Probability of Discharge to Community (discussed later in this section). Stepwise OLS regression was used to test the addition of new variables or concepts to the FIM change models. New concepts tested included premorbid social network, medical needs, premorbid function, and the quality indicators of shortness of breath at rest, pain severity, pain treatment modality, pressure ulcers, depression, and engagement in therapy. While a hierarchical model would allow for a more theoretical specification of the variables to retain, a stepwise regression model is an efficient way of selecting and dropping the addition of new variables when trying to identify the "best" model for predicting certain outcomes. All new variables were entered, including both swallowing items, (the existing swallowing item and the AHSA item) and both depression items, (the RIC-FAS and GDS).

Several of the new variables pilot tested were modified by either collapsing categories or were made into dichotomous variables in order to test them in the models. These changes are explained below:

5.7.1 Premorbid Social Network (Admission Section):

- Patient Lives With (N17)—recoded as three dummy variables, "Lives Alone," "Lives with Spouse" (reference group in the analysis), "Lives with Others."

- Assisting Persons (N18)—recoded as three dummy variables: "Patient No Assist" (reference group in the analysis), "Patient Assist in Home," (person residing in the home excluding paid help) and "Patient Assist All Others" (relatives, friends or neighbors living outside the home, paid help, none of the above).

- Type of Patient Assistance (N19)—recoded as three dummy variables: "Patient Assist Type None" (reference group in the analysis), "Patient Assist Type One (created if just one of categories 1-7 was completed affirmatively and "Patient Assist Type More than One" (created if more than one of categories 1-7 was completed affirmatively).

- Caregiver Expectations at Admission (N20)—recoded as three dummy variables: "Caregiver Self" (reference group in the analysis), "Caregiver Family," (spouse or significant other, daughter or son, other family member) and "Caregiver Other" (friend, neighbor community, church member, paid help, more than one person).

- Caregiver Frequency Expectations on Admission (N21)—recoded as two dummy variables: "Several Times (Day and/or Night)" (reference group in the analysis) and "Once Daily or Less."

5.7.2 Medical Information:

- Orientation (N27a, N27b, N27c)—Recoded as two dummy variables: "Oriented" (reference group in the analysis) refers to if the patient was oriented times three, and "Disoriented" refers patients who were disoriented in any category.

- Swallowing Measure on Admission (N28a)—recoded as two dummy variables: "No Swallowing Problem" (reference group in the analysis) and "Swallowing Problem" (all other categories)

5.7.3 FIM Instrument

- Premorbid Function—these items were created the same way as the FIM admission variables.

5.7.4 Quality Indicators:

- Pain Management (N50b)—Recoded as three dummy variables: "No Pain" (reference group in the analysis), "One Pain Modality, and "Two or More Pain Modalities" (refers to when the patient's pain is managed using one or two or more of the modalities listed on the pilot tested IRF-PAI form).

- Number of Pressure Ulcers on Admission (51a)—Recoded as two dummy variables: "No Pressure Ulcer" (reference group in the analysis) and "One or More Pressure Ulcers."

- RIC-FAS Depression Score Admission (Mood) (N52)—Recoded as four dummy variables: "No Mood Problem Admission" (reference group in the analysis), "Minimum Mood Problem Admission" (Category 6), "Moderate Mood Problem Admission" (Categories 3, 4, 5) and "Severe Mood Problem Admission" (Categories 1, 2).

- GDS (Life Empty) (N53a)—This variable refers to the first question asked on the GDS. "Do you feel that your life is empty?" It was coded as two, dummy variables: "Patient Life Empty Admission No" (reference group in the analysis) and "Patient Life Empty Admission Yes."

- Engagement (Coping) (N55)—This variable was recoded as four dummy variables: "No Coping Problem" (reference group in the analysis) and "Minimum Coping Problem Admission" (Category 6), "Moderate Coping Problem Admission" (Categories 3, 4, 5), and "Severe Coping Problem Admission" (Categories 1, 2).

5.7.5 Stepwise Regressions Models

For the five ordinary least squares regression models examining the changes in FIM scores, we began with independent variables used in the basic models, Gender, Married, Pre-Hospital Living Setting, Age, Impairment Group, and Charlson Index. Next we utilized stepwise regression to add the various concepts captured by the new variables on the pilot tested form.

This process allowed for each conceptual item in the piloted IRF-PAI form to be tested to determine which variables contributed to increasing the explanatory power of the models. The stepwise regressions added groups of variables if significant at P=0.1 level and removed variables from the models if their significance decreased to less than P=0.1 level. By adding the new pilot tested variables to the basic models, changes resulting in the adjusted R-squared values for the Pilot Sample increased for all five models, as displayed in ***Table 31***.

Table 31
Changes in R-squared value from basic OLS models after adding pilot-tested variables using stepwise regression

Outcome	Basic Model Adjusted R-Squared Value	Step-Wise Regression Model With Pilot Tested Variables R-squared Value
FIM Motor Change	0.09	0.20
FIM Cognitive Change	0.34	0.52
FIM Self-Care Change	0.19	0.30
FIM Sphincter Control Change	0.33	0.40
FIM Mobility Change	0.20	0.27

Table 32 provides details on the results of the Stepwise, OLS regressions including the significance, coefficients, and standard error of independent variables in the models.

Adding the additional pilots test variables resulted in changes in significance but not direction for the independent variables from the basic models. Gender was significant only for the FIM Sphincter Control Change model. Whether a patient was married was not significant for any model. Patients residing in a community setting prior to the admission remained positively associated with a change in the FIM score for the FIM Self-Care Change model. Age and Charlson Index score were not significant variables in any of the models.

In comparison to Replacement of Lower Extremity Joint, other impairment groups continued to be negatively associated with the change in the FIM scores (FIM Motor, Cognitive, and Self-Care). However, there were models for which an Impairment Group(s) became significant while others dropped out of the models. For example, the Impairment Groups Central Nervous System and Other became nonsignificant in the FIM Motor Change model. The Impairment Group, Spinal Cord Dysfunction, also became significant for one additional model, FIM Cognitive Change. The Impairment Group, and Stroke became insignificant for the FIM Sphincter Control Change model. The Impairment Groups Neurological and Stroke became nonsignificant in the FIM Mobility Change model.

Table 32
Stepwise OLS regressions, pilot variables, coefficients, and standard error

Dependent Variables	FIM MOTOR CHANGE		FIM COGNITIVE CHANGE		FIM SELF CARE CHANGE		FIM SPHINCTER CONTROL CHANGE		FIM MOBILITY CHANGE	
Independent Variables	Coefficient	Standard Error	Coefficient	Standard Error	Coefficient	Standard Error	Coefficient	Standard Error	Coefficient	Standard Error
Demographic Variables										
GENDER (Female)	NS	NS	NS	NS	NS	NS	1.03*	0.26	NS	NS
MARRIED (Yes)	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS
PRE-HOSPITAL LIVING SETTING (Community)	NS	NS	NS	NS	2.85*	1.72	NS	NS	NS	NS
AGE	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS
Impairment Groups										
CENTRAL NERVOUS SYSTEM	-3.03	2.50	-1.94*	0.82	-1.88	1.20	-0.05	0.68	-0.70	1.18
SPNAL CORD DYSFUNCTION	-5.43*	2.51	-1.44*	0.79	-4.03*	1.19	-1.82*	0.67	0.55	1.19
NEUROLOGIC	-4.15*	2.33	0.85	0.75	-1.98*	1.10	-0.45	0.63	-1.59	1.09
MUSCULOSKELETAL	-2.48*	1.37	-0.63	0.44	-0.94	0.65	-0.01	0.37	-1.46*	0.65
STROKE	-3.36*	1.74	-1.41*	0.60	-1.04*	0.88	-0.58	0.48	-0.85	0.83
ENDURANCE	-6.73*	2.02	-2.16*	0.66	-3.34*	0.96	-1.82*	0.54	-1.34	0.96
OTHER	-0.55	1.51	-0.80	0.50	-0.37	0.72	-0.34	0.42	0.46	0.73
Severity of Illness										
CHARLSON INDEX	NS	NS	NS	NS	NS	NS	NS	NS	-0.30*	0.16
FIM Items										
FIM SELF-CARE ADM	-0.23*	0.11	0.02	0.04	-0.53*	0.05	0.07*	0.03	0.24*	0.05
FIM COGNITIVE ADM	0.02	0.08	-0.55*	0.03	-0.01	0.04	0.05*	0.02	-0.05	0.04
FIM SPHINCTER CONTROL ADM	-0.27	0.16	0.06	0.05	0.29*	0.08	-0.61*	0.04	0.29*	0.07
FIM MOBILITY ADM	-0.31*	0.14	-0.01	0.04	0.20*	0.06	0.06*	0.03	-0.56*	0.06
Premorbid Social Network										
Lives with										
LIVE ALONE	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS
LIVES WITH OTHERS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS
Assisting Persons										
PT ASSIST IN HOME	NS	NS	0.81*	0.32	NS	NS	NS	NS	NS	NS
PT ASSIST ALL OTHERS	NS	NS	-0.29	0.31	NS	NS	NS	NS	NS	NS
Assist Type										
ASSIST TYPE ONE	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS
ASSIST TYPE MORE THAN ONE	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS
Caregiver Expectations on Admission										
CG FAMILY ADM	-2.45*	1.03	NS	NS	-0.52	0.49	-0.60*	0.27	-0.94*	0.48
CG OTHER ADM	-6.81*	2.14	NS	NS	-2.49*	1.08	-1.28*	0.56	-2.07*	1.00
Frequency of Caregiver on Admission										
ONCE DAILY OR LESS	2.75*	1.40	NS	NS	1.31*	0.66	NS	NS	NS	NS

Table 32 (continued)
Stepwise OLS regressions, pilot variables, coefficients, and standard error

Dependent Variables	FIM MOTOR CHANGE		FIM COGNITIVE CHANGE		FIM SELF CARE CHANGE		FIM SPHINCTER CONTROL CHANGE		FIM MOBILITY CHANGE	
Independent Variables	Coefficient	Standard Error	Coefficient	Standard Error	Coefficient	Standard Error	Coefficient	Standard Error	Coefficient	Standard Error
Medical Needs										
Orientation										
DISORIENTED	NS	NS	NS	NS	-1.7*	0.91	NS	NS	NS	NS
Swallowing										
SWALLOWING PROBLEM	NS	NS	-1.29*	0.42	-1.21*	0.64	NS	NS	NS	NS
Function Scores										
Premorbid										
PREMORBID SELF CARE FUNCTION	0.17	0.12	-0.06	0.04	0.08	0.06	0.06*	0.32	0.05	0.06
PREMORBID SPHINCTER CONTROL FUNCTION	-0.12	0.23	-0.17*	0.07	-0.03	0.11	0.43	0.06	-0.03	0.11
PREMORBID COGNITIVE FUNCTION	0.22*	0.12	0.11*	0.04	0.07	0.06	0.16	0.03	0.13*	0.06
PREMORBID MOBILITY FUNCTION	0.26*	0.14	0.32*	0.47	0.15*	0.07	-0.02	0.03	0.12*	0.07
Quality Indicators										
Shortness of Breath at Rest										
SHORTNESS OF BREATH AT REST (Yes)	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS
Pain Limiting Patient's Rehab										
PAIN LIMITING REHAB ADM	3.33*	1.07	NS	0.97	1.03*	0.51	0.10*	0.29	1.03*	0.50
Most Severe Pain Rating								NS		
MOST SEVERE PAIN RATING ADM	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS
Pain Treatment Modality										
ONE PAIN MODALITY ADM	NS	NS	1.19*	0.41	NS	NS	0.21	0.35	NS	NS
TWO OR MORE PAIN MODALITIES ADM	NS	NS	0.44	0.48	NS	NS	-0.56	0.41	NS	NS
Pressure Ulcer										
ONE OR MORE PRESSURE ULCERS ADM	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS
Geriatric Depression Scale										
PATIENT LIFE EMPTY ADM (Yes)	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS
RIC FAS Scale										
MOOD PROBLEM MIN ADM	NS	NS	0.85*	0.36	NS	NS	NS	NS	NS	NS
MOOD PROBLEM MOD ADM	NS	NS	0.20	0.44	NS	NS	NS	NS	NS	NS
MOOD PROBLEM SEVERE ADM	NS	NS	-2.26*	1.19	NS	NS	NS	NS	NS	NS
Engagement										
COPES PROBLEM MIN ADM	NS	NS	-1.08*	0.37	NS	NS	NS	NS	NS	NS
COPES PROBLEM MOD ADM	NS	NS	-1.76*	0.48	NS	NS	NS	NS	NS	NS
COPES PROBLEM SEVERE ADM	NS	NS	-0.27	1.18	NS	NS	NS	NS	NS	NS
R squared	0.20		0.52		0.30		0.40		0.27	

Notes follow on next page

NOTES:

N = 464

- Significant at the 0.10 level.

Impairment Group - Replacement of Lower Extremity Joint is the reference group in the analysis.
Live With - Lives Alone is the reference group in the analysis.
Assisting Persons - No Assist is in the reference group in the analysis.
Assist Type - No Assist is the reference group in the analysis.
Caregiver Expectations - Self is the reference group in the analysis.
Caregiver Frequency - Several times a day and/or night is the reference group in the analysis.
Orientation - Oriented is the reference group in the analysis.
Swallowing - No Swallowing Problem is the reference group in the analysis.
Shortness of Breath at Rest - No shortness of breath at rest is the reference group in the analysis.
Pain Limiting Patient's Rehab - No pain is the reference group in the analysis.
Pain Treatment Modality - No pain is the reference group in the analysis.
Pressure Ulcer Adm - No pressure ulcer is the reference group in the analysis.
Patient Life Empty Adm – A “no” response to patient life empty on admission is the reference group in the analysis.
RIC-FAS Scale - No mood problem is the reference group in the analysis.
Engagement - No coping problem is the reference group in the analysis.

The Charlson Index remained significant for only one of the five models, FIM Mobility Change. Some of the FIM Admission scores remained significant in all five models and did not change in direction. For example, the variable FIM Mobility score on admission remained negatively associated with a change in the FIM Motor score at discharge inferring that patients having a higher score on admission were less likely to see gains in mobility during their rehabilitation stay.

For the Premorbid Social Network items whether the patient lived with someone was not a significant variable in any model. Patients with someone in the home to assist them relative to only self for assistance experienced an increased FIM Cognitive Change score although it was not significant in any other model. The type of assistance a patient received was not significant in any model. Relative to the expectation of the patient being their own primary caregiver on admission, the expectation of having a family member or other party as a primary caregiver was negatively associated with the FIM Change score for all the models but the FIM Cognitive Change model. Finally, relative to the expectation of patients being able to receive care several times during the day and/or night, the expectation of patients being able to receive care once daily or less was associated with an increase in change score for the FIM Motor and FIM Self-Care Change models.

In the Medical Needs Section, relative to being oriented, patients who were disoriented experienced a decreased change in FIM Self-Care Change score. This item was not significant in any other model. Patients having swallowing problems were associated with lower FIM Change scores for the FIM Cognitive and Self-Care Change models.

Whether a patient was disoriented or had swallowing problems on admission was also negatively associated with a change in the Cognitive FIM score. These patients were less likely to make as great a gain in cognitive functions as patients who were not disoriented or having trouble swallowing. Swallowing problems and disorientation may be indicating additional medical severity beyond that measured by impairment group or medical condition.

Premorbid FIM Function items were significant in all the models. For the FIM Motor Change model, the Premorbid Cognitive and Mobility Function were positively associated with a gain in the FIM Motor score. FIM Sphincter Control, Cognitive, and Mobility premorbid functions were significant variables in the FIM Cognitive Change model but moved in different directions. The Premorbid Sphincter Control was negatively associated with a gain in the FIM Cognitive score while Premorbid Cognitive and Mobility function were positive in direction. For the FIM Self-Care Change model, Premorbid FIM Mobility Function was associated with an increase in the FIM Self-Care score. Premorbid Self-Care Function was positively associated with an increase in the FIM Sphinctor Control score, and finally, Premorbid Cognitive and Mobility functions were associated with a gain in the FIM Mobility score.

For the Quality Indicators Section, the variables “Pain Limiting Rehab on Adm,” “Pain Treatment Modality,” “RIC FAS Depression Scale (mood),” and “Engagement” were significant items in the various models. Shortness of Breath at Rest, Most Severe Pain Rating, Pressure Ulcers, and the GDS at Admission, were not a significant variables in any of the models.

Relative to no pain, having some pain on admission which limited the patient's ability to participate in caring for themselves was positively associated with a change in the Motor FIM, Self-Care, Sphincter Control, and Mobility Change scores. Pain Treatment Modility refers to item N50B on the revised IRF-PAI form measuring what methods for pain management were being used during the rehabilitation stay, if any. Relative to no pain treatment, One Pain Treatment Modality on Admission was associated with a positive change in the FIM Cognitive Change model.

The GDS depression item "Do you Feel Your Life is Empty?" was not significant in any of the models but the RIC FAS depression scale (mood) was significant in the FIM Cognitive Change model. A minimum mood problem was associated with an increase in the FIM Cognitive score while a severe problem was negatively associated with a change in the score. Finally, the RIC-FAS engagement item (coping) was negatively associated with a change in FIM Cognitive score for any problem (minimum, moderate and severe) relative to patients having no problem.

5.8 Preliminary Risk Adjustment Logistic Regression Models

Two sets of models were run examining the likelihood of a patient being discharged to a community setting versus any other setting. One model analyzed four combination FIM scores while the other model retained the FIM scores on an aggregate level (two scores). Model One included the variables FIM Cognitive, Sphincter Control, Mobility, and Self-Care Admission scores. Model 2 included the variables FIM Motor and Cognitive Admission scores. Both models included Gender, Marital Status, Pre-Hospital Living Setting, Age, Impairment Group, and Charlson Index. The models were run using the National, Nine Hospitals, and Pilot samples and all were significant. ***Table 33*** provides details on direction of the odds ratio when the variable was significant at the P=0.1 level.

For the National Sample, all of the demographic variables, Impairment Group variables, Charlson Index, and FIM Admission scores were significant for both sets of models. Patients who were female, married, lived in the community prior to admission, or had higher FIM Admission scores were more likely to be discharged to community. Relative to Replacement of Lower Extremity Joint, patients in all other Impairment Groups were less likely to be discharged to community, as were older patients or patients with greater degrees of comorbidity (higher Charlson Index scores). The models were then run on the Nine Hospital Sample and demonstrated the same results as the National Sample with one exception, Pre-Hospital Living Setting dropped out of Model 1 as being significant in the model.

The models were then run on the Pilot Sample and independent variables dropped out of the models as significant likely due to sample size. The demographic variables Gender, Pre-Hospital Living Setting and Age were not significant variables in either model. Whether a patient was married remained a significant variable in Model 1 denoting that a patient was more likely to be discharged to a community setting if married.

Table 33
Logistic regression: Odds of patient being discharged to community

	Model 1			Model 2		
	National	Nine Hospitals	Pilot	National	Nine Hospitals	Pilot
Independent Variables						
Demographic Variables						
GENDER (Female)	+	+	NS	+	+	NS
MARRIED (Yes)	+	+	+	+	+	NS
PRE-HOSPITAL LIVING SETTING (community)	+	NS	NS	+	+	NS
AGE	--	--	NS	--	--	NS
Impairment Groups						
CENTRAL NERVOUS SYSTEM	--	--	NS	--	--	NS
SPINAL CORD DYSFUNCTION	--	--	--	--	--	NS
NEUROLOGIC	--	--	NS	--	--	NS
MUSCULOSKELETAL	--	--	NS	--	--	NS
STROKE	--	--	NS	--	--	NS
ENDURANCE	--	--	--	--	--	--
OTHER	--	--	--	--	--	NS
Severity of Illness						
CHARLSON INDEX	--	--	--	--	--	NS
FIM Items						
FIM COGNITIVE ADM	+	+	NS	+	+	NS
FIM MOTOR ADM				+	+	+
FIM SPHINCTER CONTROL ADM	+	+	NS			
FIM MOBILITY ADM	+	+	+			
FIM SELF-CARE ADM	+	+	+			
Likelihood Ratio	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001

Notes:
( ) refer to reference category in the model.
The impairment group "Replacement of Lower Extremity Joint" is the reference group in the analysis.
FIM Adm Score:
Model 1 The scores for the following variables were entered in the model: FIM Cognitive Adm, FIM Sphincter Control Adm, FIM Mobility Adm, and FIM Self-Care Adm.
Model 2 The scores for the following variables were entered in the model: FIM cognitive Adm and FIM Motor Adm.
Significance level = 0.10

Four of the seven variables became non-significant in Model 1 while all of the Impairment Group variables but Endurance became non-significant in Model 2. However, when significant, the Impairment Group variables did not change direction. Relative to Replacement of the Lower Extremity Joint, patients in the other Impairment Groups were less likely to be discharge to a community setting. The Charlson Index remained a significant variable for Model 1 but not Model 2.

Finally, FIM Cognitive Admission scores were non-significant in either model. However, in Model 1, the FIM Mobility and Self-Care Admission scores remained positive although the Sphincter Control Admission score became non-significant. For Model 1, the FIM Motor Admission score remained significant. Thus, patients with higher FIM Admission scores were more likely to be discharged to community.

5.8.1 Logistic Regressions Testing Pilot Items

The pilot-tested variables were then added to the basic logistic models, one section of the IRF-PAI form at a time, to examine what variables contributed to the models using the Pilot Sample. ***Table 34*** provides details on the odds ratio and standard errors for the independent variables for the final step when all the pilot tested items from all the sections of the revised IRF-PAI form were added to the models. For both models, Gender remained significant and increased the likelihood of being discharged to a community setting. Marital status, Pre-Hospital Living Setting, Age, and Charlson Index score were no longer significant in the models.

For Model 1, relative to Replacement of Lower Extremity Joint, patients in the Impairment Groups Spinal Cord Dysfunction, Neurologic, Endurance and Other were less likely to be discharged to a community setting. Patients with higher FIM Sphincter Control and Mobility scores were more likely to be discharged to a community setting. For Model 2, the Impairment Groups, Central Nervous System, Spinal Cord Dysfunction, and Neurologic became significant in the model. Relative to patients in the Impairment Group Replacement of Lower Extremity Joint, Patients in these Impairment Groups were less likely to be discharged to a community setting.

In the Premorbid Social Network section of the IRF-PAI form, patients who lived alone relative to living with a spouse were less likely to be discharged to community in Model 2. The type of assistance and frequency of caregiver assistance on admission were not significant variables in either model. Relative to no assistance, patients with “other” assistance were more likely to be discharged to a community setting for both models. Consequently, patients with “other” caregiver expectations on admission were less likely to be discharged to a community setting relative to patients with “self” as the primary caregiver expectation on admission.

None of the items tested in the Medical Needs Section (Orientation, Swallowing, Shortness of Breath at Rest) were significant in the models. For Premorbid FIM Function items, only Premorbid Sphincter Control Function was significant and only in one model. Patients with a higher Premorbid Sphincter Control Function score were less likely to be discharged to a community setting for Model 2 that disaggregates the FIM items.

Table 34
Logistic regressions, pilot variables, odds ratio, and standard error

Independent Variables	Model 1 Odds Ratio	Model 1 Standard Error	Model 2 Odds Ratio	Model 2 Standard Error
Demographic Variables				
GENDER (Female)	1.96*	0.36	1.80*	0.35
MARRIED (Yes)	0.94	0.60	0.99	0.59
PRE-HOSPITAL LIVING SETTING (Community)	1.09	1.10	1.22	1.08
AGE	0.98	0.01	0.98	0.01
Severity of Illness				
CHARLSON INDEX	0.98	0.11	0.88	0.10
Impairment Groups				
CENTRAL NERVOUS SYSTEM	0.31	0.96	0.35*	0.93
SPINAL CORD DYSFUNCTION	0.21*	0.88	0.23*	0.87
NEUROLOGIC	0.14*	0.84	0.17*	0.83
MUSCULOSKELETAL	0.42	0.61	0.41	0.59
STROKE	0.78	0.77	0.85	0.75
ENDURANCE	0.14*	0.84	0.17*	0.79
OTHER	0.32*	0.69	0.38	0.66
FIM Items				
FIM SELF-CARE ADM	1.08	0.04		
FIM SPHINCTER CONTROL ADM	1.06*	0.06		
FIM MOBILITY ADM	1.11*	0.05		
FIM COGNITIVE ADM	0.99	0.03	0.87	0.03
FIM MOTOR ADM			1.08*	0.18
Premorbid Social Network				
Lives With				
LIVE ALONE	0.27	0.65	0.31*	0.63
LIVES WITH OTHERS	0.64	0.70	0.62	0.68
Assisting Persons				
PT ASSIST IN HOME	1.93	0.48	1.81	0.46
PT ASSIST ALL OTHERS	2.89*	0.45	3.05*	0.43
Assist Type				
ASSIST TYPE ONE	0.77	0.52	0.85	0.51
ASSIST TYPE MORE THAN ONE	0.59	0.52	0.61	0.50
Caregiver Expectations on Admission				
CG SPOUSE ADM	0.99	0.42	1.03	0.41
CG OTHER ADM	0.19*	0.67	0.21*	0.66
Frequency of Caregiver on Admission				
ONCE DAILY OR LESS	1.59	0.55	1.72	0.57

(continued)

Table 34 (continued)
Logistic regressions, pilot variables, odds ratio, and standard error

Independent Variables	Model 1 Odds Ratio	Model 1 Standard Error	Model 2 Odds Ratio	Model 2 Standard Error
Medical Needs				
Orientation				
DISORIENTED	0.58	0.55	0.51	0.53
Swallowing				
SWALLOWING PROBLEM	0.54	0.42	0.59	0.40
Shortness of Breath at Rest				
SHORTNESS BREATH AT REST (Yes)	1.32	0.54	1.03	0.51
Premorbid Function Score				
PREMORBID SELF-CARE FUNCTION	0.92	0.07		
PREMORBID SPHINCTER CONTROL FUNCTION	0.83*	0.10		
PREMORBID MOBILITY FUNCTION	1.08	0.05		
PREMORBID COGNITIVE FUNCTION	0.99	0.05	1.01	0.04
PREMORBID MOTOR FUNCTION			0.99	0.01
Quality Indicators				
Pain Limiting Patient's Rehab				
PAIN LIMITING REHAB ADM	0.95	0.43	1.05	0.42
Most Severe Pain Rating				
MOST SEVERE PAIN RATING ADM	0.93*	0.07	0.92	0.06
Pain Treatment Modality				
ONE PAIN MODALITY ADM	1.10	0.56	1.21	0.54
TWO OR MORE PAIN MODALITIES ADM	0.61	0.66	0.70	0.64
Pressure Ulcer ADM				
ONE OR MORE PRESSURE ULCERS ADM	1.12	0.43	1.86	0.42
Geriatric Depression Scale				
PATIENT LIFE EMPTY ADM (Yes)	0.36*	0.48	0.40*	0.48
RIC -FAS Depression Scale				
MOOD PROBLEM MIN ADM	1.32	0.42	1.50	0.42
MOOD PROBLEM MOD ADM	0.65	0.47	0.65	0.46
MOOD PROBLEM SEVERE ADM	5.42	1.14	5.99	1.10

(continued)

Table 34 (continued)
Logistic regressions, pilot variables, odds ratio, and standard error

Quality Indicators (continued)				
Engagement				
COPES PROBLEM MIN ADM	0.30*	0.46	0.40*	0.43
COPES PROBLEM MOD ADM	0.21*	0.53	0.26*	0.50
COPES PROBLEM SEVERE ADM	0.16*	1.11	0.16*	1.06

NOTES:

N = 464

* - Significant at the 0.10 level.

Impairment Group - Replacement of Lower Extremity Joint is the reference group in the analysis.

Live With - lives alone is the reference group in the analysis.

Assisting Persons - No Assist is in the reference group in the analysis.

Assist Type - No assist is the reference group in the analysis.

Caregiver Expectations - Self is the reference group in the analysis.

Caregiver Frequency - Several times a day and/or night is the reference group in the analysis.

Orientation - Oriented is the reference group in the analysis.

Swallowing - No swallowing problem is the reference group in the analysis.

Shortness of Breath at Rest - No shortness of breath at rest is the reference group in the analysis.

Pain Limiting Patient's Rehab - No pain is the reference group in the analysis.

Pain Treatment Modality - No pain is the reference group in the analysis.

Pressure Ulcer Adm - No pressure ulcer is the reference group in the analysis.

Patient Life Empty Adm – A “no” response to patient life empty on admission is the reference group in the analysis.

RIC-FAS Scale - No mood problem is the reference group in the analysis.

Engagement - No coping problem is the reference group in the analysis.

For the Quality Indicator Section of the IRF-PAI form, the first items analyzed were various concepts concerning pain. The item referring to pain limiting patient's ability to participate in the self-care process on admission was not significant in either model nor pain treatment modality on admission. Patients' most severe pain on admission was significant for Model 1. Patients with higher pain scores on admission were less likely to be discharged to a community setting. Whether a patient had a pressure ulcer on admission relative to not having a pressure ulcer on admission was not a significant variable in either model.

Next, the GDS and RIC-FAS (mood) depression items were analyzed. The first question in the GDS is "Do you feel your life is empty?" Patients who answered this item "Yes" on admission were less likely to be discharged to a community setting relative to patients answering "No" to this question. The RIC-FAS depression (mood) item was not significant in either model. Finally, the RIC-FAS engagement (copes) item was significant for both models. Relative to patients without engagement problems, patients with minimum, moderate, or severe coping problems on admission were less likely to be discharge to a community setting.

SECTION 6
RECOMMENDATIONS AND NEXT STEPS

The goal of this work is to provide CMS with recommendations regarding the types of criteria needed to monitor quality in the inpatient rehabilitation facilities (IRFs). To accomplish this, RTI has compiled a list of the TEP goals, related results from the pilot test effort, and suggested the next analytic steps, and developed a set of recommendations that also take into account the potential reporting burden for IRFs. RTI has attempted to eliminate duplicate measures and select only the best for capturing the conceptual domain being measured. We provide CMS with a parsimonious set of indicators that achieve their goal of monitoring quality while minimizing the resources required to collect data. For example, no items were included that would require every hospital to add a specific discipline (such as a psychiatrist or speech pathologist) to the treatment team for the sole purpose of collecting an item. Second, the operational aspects of each item were taken into account. Hospitals were asked to comment on both the usefulness and usability of each item in the pilot test. Having an item that met the conceptual goals but could not be scored reliably, for instance, would not contribute to a valid set of data.

This section presents our recommendations based on the results of the literature review, TEP members' input, and the analysis of the pilot test data. The section is organized into three broad parts:

- Modifications to the IRF-PAI instrument
- Empirical models and analyses
- Next steps

6.1 Modifications to the IRF-PAI instrument

The IRF-PAI instrument that was fielded in the pilot test included additions in a number of domains: pre-morbid social network, pre-morbid function, medical needs, and quality indicators. We tested multiple versions of certain items (swallowing, depression, respiratory status) to determine which worked best, if at all, in measuring IRF populations. In addition to selecting the best measure of those concepts, only those items that have contributed to the explanatory power of one of the outcomes models are included in the recommendations.

Several items were also removed from the original IRF-PAI. These include:

- Comatose and dehydration: less relevant for IRF populations;
- Clearing airway: poor measure for IRF populations;
- Balance: difficult to define and measure in a standard manner; and
- Falls: expected during a rehabilitation process. Serious injuries, which should not occur, can be identified through the ICD-9 codes, making this item unnecessary.

As shown in ***Table 35***, each of the variables discussed in the following recommendations were significantly associated with at least one outcome or were considered conceptually relevant; few were associated with every dependent variable.[9] This summary table shows the variability in the types of items that were associated with the total FIM motor change score, total FIM cognitive change score, and each of the decomposed FIM motor activities as well as the probability of discharge to the community. ***Table 36*** shows the degree to which the addition of the new items increased the explanatory power of each of these models.

Based on these results, the domains, summary of key field test findings, and recommendations are shown on ***Table 37***. To summarize, we recommend the following changes to the IRF-PAI:

- Retain the pre-morbid social network questions involving residence and caregivers—this concept was well received by both the Technical Expert Panel and the field test sites. The belief that prior social network influences discharge to a community setting is one that is recognized as relevant.

- Retain the two questions currently on the IRF-PAI that collect information on pre-hospital vocations. While these questions are not relevant to the majority of the general Medicare population, they are relevant to the high number of younger disabled beneficiaries and their expected outcomes.

- Replace delirious with the assessment of orientation to person, place, and time. This item is understandable to the IRF staff, easy to gather, and relevant to IRF outcomes.

- Retain the current item for swallowing and consider additional research prior to including ASHA's Swallowing Functional Communication Measure in the IRF-PAI.

- Include on the IRF-PAI form a pre-morbid FIM score for each item; however, collapse to three rather than seven categories.

- Keep the FIM goals item. While it should not be used to measure quality due to its intended use as a planning tool rather than an assessment measure, some hospitals find it useful internally.

- Keep two of the three pain items; keep 50A which is part of existing IRF-PAI and add 50C which assesses the impact of pain on the patient's ability to participate in the therapeutic process.

[9] This summary is based on the results presented in Section 5, Tables 32 and 34.

Table 35
Summary table of the multivariate models presented in Section 5

Independent Variables	FIM Motor Change	FIM Cognitive Change	FIM Self Care	FIM Sphincter Control Change	FIM Mobility Change	Discharge to Community 1	Discharge to Community 2
Demographic Variables							
Gender (Female)				+		+	+
Married (Yes)							
Pre-Hospital Living Setting (Community)			+				
Age							
Severity of Illness							
Charlson Index					-		
Impairment Groups							
Central Nervous System		-					-
Spinal Cord Dysfunction	-	-	-	-		-	-
Neurologic	-		-			-	-
Musculoskeletal	-				-		
Stroke	-	-	-				
Endurance	-	-	-	-		-	-
Other						-	
FIM Items							
FIM Self-Care Adm	-		-	+	+		
FIM Sphrincter Control Adm		-		+		+	
FIM Mobility Adm			+	-	+	+	
FIM Cognitive Adm	-		+	+	-		
Premorbid Social Network							
Lives With							
Lives Alone							
Lives with Others							-
Assisting Persons							
PT Assist in Home		+					
PT Assist All Others						+	+
Assist Type							
Assist Type One							
Assist Type More than One							
Caregiver Expectations on Admission							
P CG Family Adm	-			-	-		
P CG Other Adm	-		-	-	-	-	-
Frequency of Caregiver on Admission							
Once daily or less	+		+				
Medical Needs							
Disoriented (Yes)			-				
Swallowing Problems (Yes)		-	-				
Shortness of Breath at Rest (No)							

(continued)

Table 35 (continued)
Summary table of the multivariate models presented in Section 5

Independent Variables	FIM Motor Change	FIM Cognitive Change	FIM Self Care	FIM Sphincter Control Change	FIM Mobility Change	Discharge to Comm. 1	Discharge to Comm. 2
Function Scores							
Premorbid Self Care Function				+			
Premorbid Sphincter Control Function		-				-	
Premorbid Cognitive Function	+	+			+		
Premorbid Mobility Function	+	+	+		+		
Quality Indicators							
Pain Limiting Rehab Adm.	+		+	+	+		
Most Sever Pain Rating Adm.						-	
Pain Treatment Modality							
One Pain Modality Adm.		+					
Two or More Pain Modality Adm.							
Pressure Ulcer							
One or more Pressure Ulcer Adm.							
Geriatric Depression Scale							
Patient Life Empty Adm. (Yes)						-	-
RIC FAS Scale							
Mood Problem Min. Adm.		+					
Mood Problem Mod. Adm.							
Mood Problem Severe Adm.		-					
Engagement							
Copes Problem Min Adm.		-				-	-
Copes Problem Mod. Adm.		-				-	-
Copes Problem Severe Adm.						-	-

NOTES: Only those relationships that are statistically significant are included in the table.
+ = Positive statistically significant effect on dependent variable.
– = Negative statistically significant effect on dependent variable.
Impairment Groups were relative to the Replacement of Lower Extremity.

SOURCE: Section 5, Tables 32 and 34

Table 36
Explanatory Power of Models without and with the additional Pilot test items (Adjusted R-Squares)

	Existing Items		New Items
	National	Pilot	Pilot
FIM Motor Change	0.08	0.07	0.2
FIM Cognitive Change	0.24	0.32	0.52
FIM Self Care Change	0.12	0.16	0.3
FIM Sphincter Control Change	0.33	0.31	0.4
FIM Mobility Change	0.15	0.17	0.27

SOURCE: Section 5, Tables 30 and 32

- Keep the PUSH tool for pressure ulcers, but re-assess this item after collecting national data. Most hospitals document ulcers in some way. Having a measure of ulcer severity is important for identifying worsening ulcers. The other two ulcer questions should also be retained as they are important risk adjusters. Patients with ulcers will have lower participation in therapy.

- Replace the RIC-FAS depression item and the 4-item GDS with the Yale Depression Screen which is a one question item. This mood disorder screen may be a useful risk adjuster.

- Keep the RIC-FAS engagement item. Every hospital felt this was a key factor affecting outcomes and agreed that it should be considered in measuring quality of care.

- Re-align the discharge disposition question to map to the pre-morbid social network housing question.

6.2 Empirical Models and Analysis

As described in Section 5, the analyses performed on the Pilot Sample are exploratory**.** Interpretation of the results must take into consideration the small sample size, limited numbers of facilities participating in the pilot test, and the many questions and issues the hospitals raised regarding some of the piloted items being tested. Nevertheless, the outcomes of the analyses provide important direction for future work. Based on the results of the preliminary analysis, RTI's recommendations are:

- Condition-specific analyses need to be conducted using a larger sample—the limited field test precluded our ability to build risk-adjusted models that are condition specific. However, all our preliminary analyses demonstrated that the specific medical conditions leading to the rehabilitation stay significantly effected in outcomes. These additional analyses need to be performed using national data, given the small numbers of admissions for some of the impairment groups.

Table 37
Recommended Changes to the Inpatient Rehabilitation Facilities - Patient Assessment Instrument (IRF-PAI)

Conceptual Domain	Field Test Item(s)	Field Test Findings	Recommended Item(s)	Rational for Recommendation
Admission Information				
	Pre-Admission Residence (N16)	Item completion nearly 100%; 96% reported living within own residence or with family.	Pre-Admission Residence (N16)	We recommend CMS keep N16 - N19B. They are relatively easy to collect, have some variability and face validity for experts and clinical professionals in the field. Select modifications are provided.
	Patient lives with (N17)	Item completion nearly 100%; good variation with about one-half living with spouse and one-third living alone.	Patient lives with (N17)	
	Assisting Person(s) (N18)	Responses had good distribution; in only 7 cases were staff unable to assess type of assistance received premorbidly.	Assisting Person(s) (N18A) To avoid confusion, add a "none" or "not applicable" category.	
	Type of primary caregiver assistance in 3 months prior to onset (N19)	Responses ranged from ADL to IADL assistance as well as psychosocial support or financial assistance; 16% of cases had missing data. Staff reported some difficulty with clarity needed if "no care" was a response to N18.	Type of primary caregiver assistance in 3 months prior to onset (N18B)	

(continued)

Table 37 (continued)
Recommended Changes to the Inpatient Rehabilitation Facilities - Patient Assessment Instrument (IRF-PAI)

Conceptual Domain	Field Test Item(s)	Field Test Findings	Recommended Item(s)	Rational for Recommendation
Pre-morbid Social Network	Primary caregiver likely to take lead responsibility for providing and managing patient's care (N20) How often could patient receive assistance from their caregiver (N21)	One-third of sites reported some difficulty classifying patients in only one of the responses offered for N19B; some suggested number of hours received assistance rather than times per week or day; another suggested "several times during night" or "several times during day" as response options. Inadequate response options if N19A is self. A suggestion was to add skip pattern to N19A; or to add "none" category to differentiate self care from lack of caregiver.	Primary caregiver likely to take lead responsibility for providing and managing patient's care (N19A) How often could patient receive assistance from their caregiver (N19B) We are reluctant to recommend number of hours received because of recall difficulty. We do recommend adding the "none" category as well as the skip pattern for self-care. Keep pre-hospital vocational category and effort level (items 18 and 19 on original IRF-PAI).	Although these data include Medicare patients, among those who are the under 65 disabled prior employment may be a proxy for a healthier patient or one that may work harder to get back to work. For this reason it may be useful as a risk adjustment variable.

(continued)

Table 37 (continued)
Recommended Changes to the Inpatient Rehabilitation Facilities - Patient Assessment Instrument (IRF-PAI)

Conceptual Domain	Field Test Item(s)	Field Test Findings	Recommended Item(s)	Rational for Recommendation
Medical Needs Section				
Orientation	Is patient oriented to person? (N27A) Is patient oriented to place? (N27B) Is patient oriented to time? (N27C)	Little variation in orientation to person (98.3% oriented), a bit more variation in orientation to place (89.9%) and time (85.4%). Item response nearly 100%.	Is patient oriented to person? (N25A) Is patient oriented to place? (N25B) Is patient oriented to time? (N25C)	We recommend keeping the orientation items and evaluate further using larger national data set. Both TEP members and pilot hospital staff thought these factors affected outcomes. Replaces "Delirious" on the current IRF-PAI form.
Swallowing	Swallowing Functional Communication Measure (FCM) adapted from ASHA's National Outcomes Measurement System (NOMS) (N28A) Swallowing Status (3-level swallowing ability) (N28B)	Comments from field test were mixed with difficulties encountered using the 7-level ASHA item; staff view of the helpfulness of definitions were mixed; despite focus on this item, there was expressed need for additional training. The 3-level swallowing test was easy to use.	Swallowing Status (3-level swallowing ability) (Item 27)	We recommend that CMS keep the 3-level item and not include the FCM at this time. There is difficulty with data collection – categories 1 and 7 captured most of the patients. These two categories map to the dichotomy in the original item and seem to be sufficient for risk adjustment purposes. More rigorous validation of the FCM scale is warranted for use as an outcome variable prior to use in the IRF-PAI. Data collection would have implications for additional facility burden without good evidence that this level of detail is warranted.

(continued)

Table 37 (continued)
Recommended Changes to the Inpatient Rehabilitation Facilities - Patient Assessment Instrument (IRF-PAI)

Conceptual Domain	Field Test Item(s)	Field Test Findings	Recommended Item(s)	Rational for Recommendation
FIM Instrument				
Premorbid Functioning	FIM Score 3 months prior to onset (Item 39)	Data gathering is easier when family present; concern was expressed as to whether the categories could be collapsed so premorbid function is captured but not to the level of specificity required by the FIM – perhaps 3 categories rather than 7.	FIM Score 3 months prior to onset (Item 39)	We recommend keeping the item; however, collapse categories to 3 rather than 7, including ability to perform independently, requiring supervision/assistance, and unable to perform.
FIM Goals	FIM Score Goals	Most hospitals are tracking FIM goals on the patient medical record. While this item should not be used to measure quality because of its intended use for internal planning, it could be retained on the form for internal operational purposes.		
Quality Indicators				
Pain	Rate the highest level of pain reported by the patient within the assessment period (50A) Methods of pain management used during the stay (N50B) Does pain limit the patient's ability to participate in the therapeutic process (N50C)	Good item response and adequate distribution; assessment is challenging with cognitively impaired patients; there was a recommendation to add whether pain also interfered with participation in therapy; some of the facilities would have liked more than a dichotomous response to N50C to assess whether pain limits participation all the time, part of the time, or never.	Rate the highest level of pain reported by the patient within the assessment period (Item 51) Does pain limit the patient's ability to participate in the therapeutic process (N51C): Modify the response category to capture pain interfering with participation in the *therapeutic* process.	We recommend keeping the pain items to use primarily as a risk adjustment item; consider as an outcome variable with additional analyses of national data. 50A: Hospitals are familiar with using some variant of this scale internally. N50B: Eliminate method of pain management as it is not useful for risk adjustment and is not an outcome variable.

(continued)

Table 37 (continued)
Recommended Changes to the Inpatient Rehabilitation Facilities - Patient Assessment Instrument (IRF-PAI)

Conceptual Domain	Field Test Item(s)	Field Test Findings	Recommended Item(s)	Rational for Recommendation
Pressure Ulcers/ PUSH Tool	Number of current pressure ulcers (Item 51A) Highest current pressure ulcer stage (Item 51B) PUSH Tool v. 3.0 (Items 51C – 51F)	At admission 85% had no ulcers and at discharge 87.2% had no ulcers. The presence of an ulcer is intended to be used as a risk adjuster.	Number of current pressure ulcers (Item 52B) Highest current pressure ulcer stage (Item 52A) PUSH Tool v. 3.0 (Items 52C – 52F)	The wound care specialists at the participating hospitals found these items to be valuable, and many hospitals were already collecting this type of information, but not necessarily reporting it to CMS. Because this is not currently used as an outcome variable, it might be sufficient to note the presence of an ulcer at admission and not require the entire PUSH score. Alternatively, improvement in ulcers may be considered as an outcome variable, in which case documenting the PUSH score at admission and at discharge may be valuable. It was noted that these items track multiple ulcers (worst ulcer at the time of assessment) rather than changes in one ulcer. National data would be needed to evaluate the utility of this potential outcome.

(continued)

Table 37 (continued)
Recommended Changes to the Inpatient Rehabilitation Facilities - Patient Assessment Instrument (IRF-PAI)

Conceptual Domain	Field Test Item(s)	Field Test Findings	Recommended Item(s)	Rational for Recommendation
Mood/ Depression	RIC-FAS Mood/Depression Item (N52) 4-item Geriatric Depression Scale (N53A-N53D) Can the patient answer the prior questions? (N53F) Was the patient referred to a mental health professional for assessment or for any reason? (N54)	Two depression items were tested and both were found wanting. GDS had some problems with high rates of missing data and some questionable skips. Pilot study participants were generally negative about this item and expressed difficulty with scoring it for cognitively impaired patients. Staff at 4 of the 9 sites felt uncomfortable asking the GDS items, particularly the screening items. Several asked whether a psychiatrist would be needed to assess this item; felt they were not the appropriate discipline to "diagnose" a problem, and pointed out that many smaller hospitals do not have a psychiatrist on staff except when needed. RIC-FAS depression was reported to be easier to administer than the GDS; however, comments included that it was too open to interpretation.	Yale Depression Screen: Do you often feel sad or depressed? (N55)	Pilot study participants were generally negative about both the RIC-FAS and GDS mood/ depression items and worried about their ability to score the items correctly and the availability of hospital resources necessary to assess depression and mood problems. The recommended Yale Depression Screen item was suggested by the psychologists on the TEP as being valid in this population, easy to collect, and targeted to patient's current affective moods which are the factor that would affect their ability to participate in the therapeutic process.

(continued)

Table 37 (continued)
Recommended Changes to the Inpatient Rehabilitation Facilities - Patient Assessment Instrument (IRF-PAI)

Conceptual Domain	Field Test Item(s)	Field Test Findings	Recommended Item(s)	Rational for Recommendation
Engagement	RIC FAS Engagement Item (N55)	The engagement item was received quite well by the sites, with comments that suggested this was an important concept to the clinicians at IRFs. There was some concern expressed about the difficulty of gathering this rating for patients with cognitive impairments. However, there was unanimous consensus that this measure is one of the most important factors affecting outcomes, and participants were pleased that it might be recognized by CMS.	RIC FAS Engagement Item (N56)	We recommend that the RIC-FAS Engagement Item be used as an indication of willingness/ability to participate in therapy. There are some challenges to using this item, so that adequate instruction and training are needed. The utility of assessing engagement is to adjust for its impact on participation in therapy. While this may be somewhat correlated with depression, these are independent concepts as demonstrated by the analysis in Section 4. Many factors besides depression may affect engagement levels. Analysis based on a larger sample is needed.

- Repeat analysis on a national level using a theory-driven technique, such as hierarchical modeling, where variables measuring domains are entered one at a time and their impact on the model is examined. The limited field test provided insufficient cases to examine the items within conditions. The utility and importance of these items will differ by condition and should be modeled separately. The analysis of field test data, by necessity, aggregated across hospitals and across conditions so that models need to be replicated using a larger pool of patients. Additionally, many of the pilot-tested variables were either aggregated into dichotomous variables or collapsed into fewer categories in order to test their effect in the models due to lack of sample size. Thus, larger samples will provide an opportunity to analyze the influence of multi-category variables more fully.

- We considered two major outcomes: discharge disposition and change in various FIM scores. Both cognitive and motor FIM scores were considered. Another method that can be used to conceptualize the outcome is to consider observed versus expected, condition-specific, risk-adjusted change scores (Stineman et al., 199X). Using this method would allow a hospital to track its performance relative to its case mix. We recommend that CMS undertake this type of analysis using national data.

6.3 Next Steps

This study was described at several national rehabilitation meetings this past year, either as part of a larger presentation on CMS' quality initiatives (Scott, 2004) or to professional rehabilitation research or provider audiences (Constantine, 2004; Gage, 2004; Deutsch, 2004). Interest was expressed in having national benchmarks that could be used to study individual hospital quality. Currently, there are several initiatives trying to use UDS^{MR} data, but not all hospitals participate in that data collection and the efforts are proprietary. Hospitals use these reports and develop other internal efforts to monitor quality for internal purposes. Where relevant, we built on the CARF-accreditation standards.

Much work remains to be done over the next few years to test these measures on the entire range of Medicare populations. Specifically, we recommend that CMS:

- Require hospital submission of the revised and recommended quality indicators. As noted in the current IRF-PAI analysis, few hospitals report voluntary items.
- Update IRFPAI training manual to include training materials on the new items. Also update the vignettes used in the training sessions to incorporate changes from the field test. Incorporate these changes into the existing training system, helpdesk, website and other resources.
- After at least one year of data:
  - Repeat and expand the data analyses using the techniques recommended above to examine importance of added items to risk adjusting outcomes.
  - Reassess level of burden on facilities relative to importance of data collected.
  - Reconvene TEP to address results based on national dataset.

REFERENCES

AGS Panel on Chronic Pain in Older Persons: The management of chronic pain in older persons. *Journal of Amer Geriatric Soc*, 46(5):635-651, 1998.

Bernard, S., E. Root, and J. O'Keefe, *Developing Quality Indicators for Inpatient Rehabilitation Facilities, Technical Expert Panel Summary, August 2004*. RTI International, Health, Social, and Economics Research, Research Triangle Park, NC, CMS Contract No. 500-00-0024-004.

Carter G, Relles D, Buchanan J, Bean D, Donyo T, Rosenfeld K, Rubinstein L, and Spritzer K. A classification system for inpatient medical rehabilitation patients: A review and proposed revisions to the functional independence measure-function related groups. A project memorandum prepared for the Health Care Financing Administration. Santa Monica, CA: RAND, 1997.

Carter G, Relles D, Buchanan J, Donyo T, Inkelas M, and Spritzer K. A prospective payment system for inpatient rehabilitation. A project memorandum prepared for the Health Care Financing Administration. Santa Monica, CA: RAND, 1997.

Charlson Index. Last updated: March 5, 2001. http/www.umanitoba.ca/centres/mchp/concept/dict/charlson.index.html. 2/19/2004

Constantine, R. et al., "Developing Quality Indicators for Inpatient Rehabilitation Facilities," presented at annual meeting of American Medical Rehabilitation Providers Association, Fall 2004.

Dobrez, G., A. Lo Sasso, and A. Heinemann, The effect of prospective payment on rehabilitative care, *Arch Phys Med Rehabil*, Vol. 85:1909-14, Dec. 2004.

Dodds, TA, Martin DP, Stolov WC, and Deyo RA, A validation of the functional independence measurement and its performance among rehabilitation in-patients. Arch Phys Med Rehabil 1993;74:531-6.

Gage, B., D. Osber, and W. Bartosch, W. Preliminary analysis from "Evaluation of BBA impact on Medicare delivery and utilization of inpatient and outpatient therapy services," (in progress), CMS Contract No. 500-00-0030.

Gage, B. et al. "Developing Quality Indicators for Inpatient Rehabilitation Facilities," presented at annual conference of California Rehabilitation Association, Western Alliance for Rehabilitation, Fall 2004.

Gillen, R., Eberhardt, T.L., and Tennen, H., Screening for depression in stroke: relationship to rehabilitation efficiency. *Journal of Stroke and Cerebrovascular Diseases* 8(5):300-306, 1999.

Granger, CV, Kelly-Hayes, M, Johnston M, Deutsch A, Braun S, and Fieldler RC, Quality and outcome measures for medical rehabilitation. In: Brandom RL, editor. Physical Medicine and Rehabilitation. Phila.: Saunders, 1995;239-53.

Hamilton, BB, Granger, CV, Sherwin FS, Zielezny M, and Tashman JS, A uniform national data system for medical rehabilitation. In: Rehabilitation Outcomes: Analysis and Measurement. Fuhrer, MJ, Ed., Paul H. Brookes, Baltimore MD, 1987.

Hamilton BB, Laughlin JA, Fiedler RC, and Granger CV, Interrater reliability of the 7-level functional independence measure (FIM). Scan. J. Rehab. Med., 26. 115-119, 1994.

Keller S, L. Greenwald, and J O'Keefe, *Developing Quality Indicators for Inpatient Rehabilitation Facilities, Draft Report on First Meeting of the Technical Expert Panel, February 20, 2002,* Washington, DC. Health, Social and Economic Research Unit, RTI International, Research Triangle Park, NC, CMS Contract No. 500-00-0024-004.

Linacre, JM, Heinemann AW, Wright, BD, Granger CV, and Hamilton BB, The structure and stability of the functional independence measure, Arch Phys Med Rehabil, 75(2), 127-32, 1994.

Quan, Hude, GA Parsons, and WA Ghali. "Validity of Information on Comorbidity Derived from ICD-9-CCM Administrative Data." Medical Care. 40(8):675-685, .

Resnick, B. and Daily M.P., The effect of cognitive status on outcomes following rehabilitation. *Family Medicine: Special Series: Evaluation Functional Status in Older Persons* 29(6):400-405, June 1997.

RIC Functional Assessment Scale, Version V, Rehabilitation Institute of Chicago, 1998.

Roos Leslie L, L Stranc, RC James, and J Li. "Complications, Comorbidities, and Mortality: Improving Classifications and Prediction." Health Services Research. 32(2):229-238. June 1997.

Scott, J., "Quality Measures for Inpatient Rehabilitation Facilities," presented at annual meeting of Association of Rehabilitation Nurses, Fall 2004.

Stineman Margaret G. "Function-Related Groups 101: A Primer." Critical Reviews™ in Physical and Rehabilitation Medicine. 10:319-358, 1998.

Stineman Margaret G., RN Ross, R Fiedler, and CV Granger. "Functional Independence Staging: Conceptual Foundation, Face Validity, and Empirical Derivation." Arch Phys Med Rehabil. 84:29-37, January 2003.

Stineman MG, Tassoni CJ, Escarce JJ, Goin JE, Granger CV, Fieldler RC, and Williams SV. Development of FRGs 2.0. A classification system for medical rehabilitation. Health Serv. Res., 32(4), 529-48, 1997.

Stineman MG, Escarce JJ, Goin JE, Hamilton BB, Granger CV, Williams SV. A case mix classification system for medical rehabilitation. Med Care. 1994;32:366-79.

Stineman MG, Shea, JA, Jette A, Tassoni CJ, Ottenbacher KJ, Fieldler R, and Granger C, The Functional Independence Measure: tests of scaling assumptions, structure, and reliability across 20 diverse impairment categories. Arch Phys Med Rehabil, 77(11), 1101-8, 1996.

Stineman MG, Goin JE, Hamilton BB, and Granger CV. Efficiency pattern analysis for medical rehabilitation. Am J Med Quality, 1995; 10(4):190-9.

Stineman MG and Granger CV. Outcome, efficiency, and time-trend analysis for stroke rehabilitation. Am J Phy Med Rehabil, 1998; 77(3):192-201.

Transcript from Technical Expert Panel Meeting, February 20, 2002.

Transcript of Technical Expert Panel Meeting, Development of Quality Indicators for Inpatient Rehabilitation Facilities, November 6-7, 2003.