April 2014

Draft Specifications for the Functional Status Quality Measures for Inpatient Rehabilitation Facilities (Version 2)

Prepared for

Center for Clinical Standards and Quality
Centers for Medicare & Medicaid Services
7500 Security Boulevard
Baltimore, MD 21244-1850

CMS Contract No. HHSM-500-2008-00021I

RTI INTERNATIONAL

[This page intentionally left blank.]

DRAFT SPECIFICATIONS FOR THE FUNCTIONAL STATUS QUALITY MEASURES FOR INPATIENT REHABILITATION FACILITIES (VERSION 2)

RTI International

CMS Contract No. HHSM-500-2008-00021I

April 2014

This project was funded by the Centers for Medicare & Medicaid Services under contract no. HHSM-500-2008-00021I. 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 International is a trade name of Research Triangle Institute.

[This page intentionally left blank.]

CONTENTS

Function Definition:

The World Health Organization's International Classification of Functioning, Disability and Health (ICF) describes the term "function" as an umbrella term that encompasses all body structures and functions, activities, and participation in daily life.[1] Examples of functioning within the components of body structures and functions include swallowing and bladder and bowel continence. Examples of functioning status within the area of activities include eating, bathing, and dressing; in the area of participation, examples include working and participating in recreational activities. As noted above, functioning is a broad term that covers various components and several levels (e.g., body, person, society).

[This page intentionally left blank.]

SECTION 1
BACKGROUND

This document describes draft specifications for four functional status quality measures for inpatient rehabilitation facilities (IRFs). This work builds on previous work, including the Development and Testing of the Continuity Assessment Record and Evaluation (CARE),[2,3] the Post-Acute Care Payment Reform Demonstration (PAC PRD),[4-8] and the Analysis of Crosscutting Medicare Functional Status Quality Metrics Using the Continuity Assessment Record and Evaluation.[9] A Technical Expert Panel (TEP) convened by RTI International was consulted during the development of these measure specifications via one in-person meeting and several conference calls.[10]

During an inpatient rehabilitation stay, treatment goals include fostering the patient's ability to manage his or her daily activities so that the patient can complete self-care and mobility activities as independently as possible and if feasible, return to a safe, active, and productive life in a community-based setting.

Given that the primary goal of rehabilitation is improvement in function, IRF clinicians have traditionally assessed and documented patients' functional status at admission and at discharge to evaluate the effectiveness of the rehabilitation care provided to individual patients, as well as the effectiveness of the rehabilitation unit or hospital overall.

In describing the importance of functional status, the National Committee on Vital and Health Statistics, Subcommittee on Health,[11] noted,

> "Information on functional status is becoming increasingly essential for fostering healthy people and a healthy population. Achieving optimal health and well-being for Americans requires an understanding across the life span of the effects of people's health conditions on their ability to do basic activities and participate in life situations, in other words, their functional status."

Research has shown differences in IRF patients' functional outcomes by geographic region, insurance type, and race/ethnicity after adjusting for key patient demographic characteristics and admission clinical status, which supports the need to monitor IRF patients' functional outcomes. For example, Reistetter[12] examined discharge motor function and functional gain among IRF patients with stroke and found statistically significant differences in functional outcomes by U.S. geographic region, insurance type, and race ethnicity group after risk adjustment. O'Brien[13] found differences in functional outcomes across race/ethnicity groups in her analysis of Medicare assessment data for patients with stroke after risk adjustment. O'Brien[13] also noted that the overall IRF length of stay decreased 1.8 days between 2002 and 2007, and that shorter IRF stays were associated with lower function at discharge.

Rehabilitation programs have traditionally conceptualized functional status in terms of the need for assistance from another person. This is the conceptual basis for the FIM® instrument (used in IRFs); the Minimum Data Set (MDS) function items (used in nursing

---

FIM® is a trademark of Uniform Data System for Medical Rehabilitation, a division of UB Foundation Activities, Inc.

homes), and the Outcome and Assessment Information Set (OASIS) function items (used in home health). In a patient-centered health care system, there is a need for standardized terminology and assessment items because patients often receive care from more than one provider. The use of standardized items and terminology facilitates clinicians speaking a common language that can be understood across clinical disciplines and practice settings.

The functional assessment items used to calculate the four IRF function quality measures are from the CARE Item Set, which was designed to standardize assessment of patients' status across acute and post-acute settings, including IRFs, long-term care hospitals (LTCHs), skilled nursing facilities (SNFs), and home health agencies (HHAs). The CARE Item Set was developed and tested as part of the PAC PRD. The functional status items (also known as data elements) on the CARE Item Set include daily activities that clinicians typically assess at the time of admission and/or at discharge to determine patient needs, evaluate patient progress, and prepare patients and families for a transition to home or another setting.

The development of the CARE Item Set and a description and rationale for each item is described in a report titled *The Development and Testing of the Continuity Assessment Record and Evaluation (CARE) Item Set: Final Report on the Development of the CARE Item Set: Volume 1 of 3*.[2]

Results of the reliability and validity testing conducted as part of the PAC PRD found the functional status items to have acceptable reliability and validity in the acute and post-acute patient populations. A description of the testing methodology and results are available in several reports, including *The Development and Testing of the Continuity Assessment Record and Evaluation (CARE) Item Set*. Volume 2 of the series is the *Final Report On Reliability Testing*[8] and Volume 3 is the *Final Report on CARE Item Set and Current Assessment Comparisons*.[3] These reports are available at http://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/Post-Acute-Care-Quality-Initiatives/CARE-Item-Set-and-B-CARE.html. A summary of the reliability and validity testing of the CARE functional status items is provided in ***Appendix A*** of this document.

The quality measures described in this document focus on self-care and mobility activities. We recognize that inpatient rehabilitation programs focus on recovery across many areas of function at the level of body structure and function, activities, and participation; however, additional research is needed to develop quality measures for other areas of functioning.

SECTION 2
QUALITY MEASURES

2.1 Quality Measure: IRF Functional Outcome Measure: Change in Self-Care Score for Medical Rehabilitation Patients

2.1.1 Summary Description

This measure estimates the average risk-adjusted mean change in *self-care function* between admission and discharge for patients discharged from IRFs.

2.1.2 Purpose/Rationale for Quality Measure

Given that the primary goal of rehabilitation is improvement in functional status, IRF clinicians have traditionally assessed and documented patients' functional status at admission and at discharge to evaluate the effectiveness of the rehabilitation care provided to individual patients, as well as the effectiveness of the rehabilitation unit or hospital overall.

Studies have shown differences in IRF patients' functional outcomes by geographic region, insurance type, and race/ethnicity after adjusting for key patient demographic characteristics and admission clinical status, which supports the need to monitor IRF patients' functional outcomes.

As stated prior, the quality measures described in this document focus on self-care and mobility activities. We recognize that inpatient rehabilitation programs focus on recovery across many areas of function at the level of body structure and function, activities, and participation. Additional research is needed to develop quality measures for other areas of functioning.

2.1.3 Population

Inclusion Criteria

The population for this quality measure includes IRF patients who are 21 years of age or older.

Exclusion Criteria

Four exclusion criteria apply to the change in self-care function score measure:

1) Patients with incomplete stays: It can be challenging to gather accurate discharge functional status data for patients who experience incomplete stays. Patients with incomplete stays include patients who are unexpectedly discharged to an acute-care setting (inpatient prospective payment system [IPPS], critical access hospital [CAH], inpatient psychiatric facility [IPF], or LTCH) because of a medical emergency; patients who die; patients who leave an IRF against medical advice; and patients with a length of stay less than 3 days.

2) Patients who are independent with CARE self-care activities at the time of admission: Patients who are independent with the CARE self-care items at the time of admission

are assigned the highest score on all the self-care items, and thus, would not be able to show functional improvement on this same set of items at discharge.

3) Patients in coma, persistent vegetative state, complete teraplegia, and locked-in syndrome are excluded, because they may have limited or less predictable self-care improvement.

4) Patients younger than age 21.

2.1.4 Functional Status Items

One important consideration about measuring functional status is that certain functional status activities may not be relevant or feasible to collect for all patients in all types of settings. For example, an IRF patient may have difficulty swallowing and may be unable to eat by mouth at the time of admission. Therefore, clinicians may indicate that a functional activity did not occur based on a patient's clinical status, because it is not safe or feasible for the patient to perform the activity.

In a recent TEP meeting convened by RTI,[10] participants were asked to describe functional activities assessed by clinicians in their own facilities as well as functional activities assessed by clinicians working in rehabilitation facilities that apply best practices. Participants noted that some of the more challenging activities are not assessed at admission but are important to assess at discharge, especially for patients returning to a community-based setting.

For the quality measure, the following functional activities are assessed and rated at the time of admission and at discharge:

Self-Care Items

Eating: The ability to use suitable utensils to bring food to the mouth and swallow food once the meal is presented on a table/tray. Includes modified food consistency.

Oral hygiene: The ability to use suitable items to clean teeth.

Toilet hygiene: The ability to maintain perineal hygiene; ability to adjust clothes before and after using toilet, commode, bedpan, or urinal.

Shower/bathe self: The ability to bathe self in shower or tub, including washing, rinsing, and drying self. Does not include transferring in/out of tub/shower.

Upper body dressing: The ability to put on and remove shirt or pajama top; includes buttoning, if applicable.

Lower body dressing: The ability to dress and undress below the waist, including fasteners; does not include footwear.

Putting on/taking off footwear: The ability to put on and take off socks and shoes or other footwear that is appropriate for safe mobility.

Self-Care Rating Scale: Codes and Code Definitions

6. **Independent** – Patient completes the activity by himself/herself with no assistance from a helper.
5. **Setup or clean-up assistance** – Helper SETS UP or CLEANS UP; patient completes activity. Helper assists only prior to or following the activity.
4. **Supervision or touching assistance** – Helper provides VERBAL CUES or TOUCHING/ STEADYING assistance as patient completes activity. Assistance may be provided throughout the activity or intermittently.
3. **Partial/moderate assistance** – Helper does LESS THAN HALF the effort. Helper lifts, holds, or supports patient's trunk or limbs, but provides less than half the effort.
2. **Substantial/maximal assistance** – Helper does MORE THAN HALF the effort. Helper lifts or holds patient's trunk or limbs and provides more than half the effort.
1. **Dependent** – Helper does ALL of the effort. Patient does none of the effort to complete the task. Or, the assistance of 2 or more helpers is required for the patient to complete the activity.

If the activity did not occur, code one of the following:

88. Not attempted due to medical condition or safety concerns

09. Not applicable

07. Patient refused

99. Not a patient goal (use at discharge only)

2.1.5 Quality Measure Calculation

The following steps are used to calculate the measure:

1) Sum the scores of the admission self-care items to create an admission self-care score for each patient, after 'activity did not occur' values are recoded. (range: 7 to 42).

2) Sum the scores of the discharge self-care items to create a discharge self-care score for each patient, after 'activity did not occur' values are recoded. (range: 7 to 42).

3) Calculate the difference between the admission self-care score and the discharge self-care score for each patient to create a change in self-care score for each patient.

4) Calculate a predicted change in self-care score for each patient using a statistical model that estimates both the average predictive effect of the patient characteristics across all IRFs and the degree to which each IRF has an effect on functional change that differs from that of the average IRF.

5) Calculate an expected change in self-care score for each patient using a statistical model that estimates the average predictive effect of the patient characteristics across all IRFs on the basis of each patient's admission characteristics (risk adjustors).

6) Calculate an average predicted change in self-care score for each IRF. This is the IRF's predicted change in self-care score.

7) Calculate an average expected change in self-care score for each IRF. This is the IRF's expected change in self-care score.

8) For each IRF, divide the IRF's predicted change score by the IRF's expected change score to create a predicted to expected ratio, and then multiply each IRF's ratio by the national average change in self-care function.

2.1.6 Risk Adjustment

Patients treated in IRFs vary in terms of primary diagnosis (i.e., impairment group), demographic characteristics, and co-existing conditions. Patients may also have different expected improvement in function on the basis of these factors. Therefore, this outcome measure is risk adjusted. Risk adjustment controls for specific patient characteristics (e.g., age or diagnosis) that may affect patients' outcomes when comparing facilities.

An initial, extensive set of risk adjustment variables was selected for this quality measure on the basis of a review of the literature and empirical findings from the PAC PRD analyses[7] as well as input from the TEP convened by RTI.[10] Using this initial set of risk adjustment variables, we have been conducting regression analyses using the PAC PRD data to help identify the best set of risk adjustors on the basis of regression coefficients, statistical significance, sample sizes, and other indicators. We also requested input on suggested risk adjustors as part of the public comment process, and are incorporating suggestions we received as we refine our risk adjustment models. Data on the reliability of CARE variables used for risk adjustment can be found in the report titled *The Development and Testing of the Continuity Assessment Record and Evaluation (CARE) Item Set: Final Report on Reliability Testing: Volume 2 of 3*.[8] The risk adjustment modeling approach follows the guidelines described by Ash et al. in the paper titled "Statistical Issues In Assessing Hospital Performance."[14]

The current list of risk adjustment variables is outlined below. This list will be updated, as appropriate, on the basis of further analyses.

The risk adjustors used for this quality measure are the following:

- **Age group at IRF admission**
  - Younger than 35 years
  - 35 to 44 years
  - 45 to 54 years
  - 55 to 64 years
  - 65 to 74 years
  - 75 to 84 years 85 years of age or older (reference category).

- **Primary rehabilitation diagnosis**
  - Stroke (reference category)
  - Brain dysfunction
  - Spinal cord dysfunction
  - Other neurological conditions
  - Fractures and other multiple trauma
  - Hip and knee replacement
  - Other orthopedic conditions (amputation, arthritis, hip fracture)
  - Cardiac conditions, pulmonary conditions, and debility
- **Medical vs. surgical acute-care diagnosis**
- **Functional status before current illness/injury: self-care**
  - Dependent
  - Some help
  - Independent, or unknown (reference category)
- **Functional status before current illness/injury: indoor ambulation**
  - Dependent or some help
  - Independent, or unknown (reference category)
- **Wheelchair use before current illness/injury**
  - Yes
  - No, or unknown (reference category)
- **Presence of severe pressure ulcer at admission** (Stage 3, Stage 4, or Unstageable pressure ulcer)
- **Cognitive abilities: Brief Interview for Mental Status (BIMS) score**
  - Severely impaired
  - Moderately impaired
  - Intact or borderline (reference category)
- **Bladder incontinence**
  - Always incontinent
  - Less than daily or daily incontinence
  - Continent or stress incontinence only or no urine output (reference category)

- **Communication: Understanding verbal content *and* expression of ideas and wants**
  - Moderate to severe communication limitations: Rarely/never understands; or sometimes understands; or rarely/never expresses self; or speech is very difficult to understand; or frequently exhibits difficulty with expression
  - Mild to no communication limitations: Usually understands or understands; or some difficulty with expression; or expression without difficulty; or unable to assess or unknown (reference category)
- **Baseline self-care function score and baseline self-care function score squared**
- **Time from illness/injury onset to rehabilitation admission**
- **Comorbidities** (hierarchical condition categories) e.g., chronic kidney disease and dialysis; metastatic cancer and acute leukemia; diabetes; delirium and encephalopathy; paraplegia; multiple sclerosis; Parkinson's and Huntington's disease; chronic ischemic heart disease

2.2 Quality Measure: IRF Functional Outcome Measure: Change in Mobility Score for Medical Rehabilitation Patients

2.2.1 Summary Description

This quality measure estimates the average risk-adjusted mean change in *mobility function* between admission and discharge for patients discharged from an IRF.

2.2.2 Purpose/Rationale for Quality Measure

Given that the primary goal of rehabilitation is improvement in function, IRF clinicians have traditionally assessed and documented patients' functional status at admission and discharge to evaluate the effectiveness of the rehabilitation care provided to individual patients, as well as the effectiveness of the rehabilitation unit or hospital overall.

Studies have shown differences in IRF patients' functional outcomes by geographic region, insurance type, and race/ethnicity after adjusting for key patient demographic characteristics and admission clinical status, which supports the need to monitor IRF patients' functional outcomes.

As stated prior, the quality measures described in this document focus on self-care and mobility activities. We recognize that inpatient rehabilitation programs focus on recovery across many areas of function at the level of body structure and function, activities, and participation. Additional research is needed to develop quality measures for other areas of functioning.

2.2.3 Population

Inclusion Criteria

The population for this measure includes IRF patients who are 21 years old or older.

Exclusion Criteria

Three exclusion criteria apply to the change in mobility function score quality measure:

1) Patients with incomplete stays: It can be challenging to gather accurate discharge functional status data for patients who experience incomplete stays. Patients with incomplete stays include patients who are unexpectedly discharged to an acute care setting (IPPS, CAH, IPF, or LTCH) because of a medical emergency; patients who die or leave an IRF against medical advice;, and patients with a length of stay less than 3 days.

2) Patients who are independent with CARE mobility activities at the time of admission: Patients who are independent with the CARE mobility items at the time of admission are assigned the highest score on all the mobility items, and thus, would not be able to show functional improvement on this same set of items at discharge.

3) Patients in coma, persistent vegetative state, complete teraplegia, and locked-in syndrome are excluded, because they may have limited or less predictable self-care improvement.

4) Patients younger than age 21.

2.2.4 Functional Status Items

One important consideration about measuring functional status is that certain functional status activities may not be safe, relevant, or feasible to collect for all patients in all types of settings. For example, an IRF patient may be unable to walk up and down stairs at the time of admission. Therefore, clinicians may indicate that a functional activity did not occur for patients on the basis of their clinical status because it is not safe or feasible for the patient to perform the activity.

In a recent TEP meeting convened by RTI, participants were asked to describe functional activities assessed by clinicians in their own facilities, as well as functional activities assessed by clinicians working in rehabilitation facilities that apply best practices. Participants noted that some of the more challenging activities are not assessed at admission, but are important to assess at discharge, especially for patients returning to a community-based setting.

For the quality measure, the following functional activities are assessed and rated at the time of admission and discharge:

Mobility Items

Roll left and right: The ability to roll from lying on back to left and right side, and roll back to back.

Sit to lying: The ability to move from sitting on side of bed to lying flat on the bed.

Lying to sitting on side of bed: The ability to safely move from lying on the back to sitting on the side of the bed with feet flat on the floor, no back support.

Sit to stand: The ability to safely come to a standing position from a position of sitting in a chair or on the side of the bed.

Chair/bed-to-chair transfer: The ability to safely transfer to and from a chair (or wheelchair).

Toilet transfer: The ability to safely get on and off a toilet or commode.

Car transfer: The ability to transfer in and out of a car or van on the passenger side. Does not include the ability to open/close door or fasten seat belt.

For patients who are walking, complete the following items:

Walk 10 feet: Once standing, the ability to walk at least 10 feet (3 meters) in room, corridor, or similar space.

Walk 50 feet with two turns: Once standing, the ability to walk 50 feet and make two turns.

Walk 150 feet: Once standing, the ability to walk at least150 feet (45 meters) in corridor or similar space.

Walking 10 feet on uneven surfaces: The ability to walk 10 feet on uneven or sloping surfaces, such as grass or gravel.

1 step (curb): The ability to step over a curb or up and down one step

4 steps: The ability to go up and down four steps with or without a rail.

12 steps: The ability to go up and down 12 steps with or without a rail.

Picking up object: The ability to bend/stoop from a standing position to pick up a small object, such as a spoon, from the floor.

Mobility Rating Scale: Codes and Code Definitions

6. **Independent** – Patient completes the activity on his/her own with no assistance from a helper.
5. **Setup or clean-up assistance** – Helper SETS UP or CLEANS UP; patient completes activity. Helper assists only prior to or following the activity.
4. **Supervision or touching assistance** –Helper provides VERBAL CUES or TOUCHING/ STEADYING assistance as patient completes activity. Assistance may be provided throughout the activity or intermittently.
3. **Partial/moderate assistance** – Helper does LESS THAN HALF the effort. Helper lifts, holds, or supports patient's trunk or limbs, but provides less than half the effort.
2. **Substantial/maximal assistance** – Helper does MORE THAN HALF the effort. Helper lifts or holds patient's trunk or limbs and provides more than half the effort.
1. **Dependent** – Helper does ALL of the effort. Patient does none of the effort to complete the task. Or, the assistance of 2 or more helpers is required for the patient to complete the activity.

If activity did not occur, code one of the following:

88. Not attempted due to medical condition or safety concerns

09. Not applicable

07. Patient refused

99. Not a patient goal (use at discharge only)

2.2.5 Quality Measure Calculation

The following steps are used to calculate the measure:

1) Sum the scores of the admission mobility items to create an admission mobility score for each patient, after 'activity did not occur' values are recoded. (range: 15 to 90)

2) Sum the scores of the discharge mobility items to create a discharge mobility score for each patient, after' activity did not occur' values are recoded. (range: 15 to 90)

3) Calculate the difference between the admission mobility score and the discharge mobility score for each patient to create a change in mobility score for each patient.

4) Calculate a predicted change in mobility score for each patient using a statistical model that estimates both the average predictive effect of the patient characteristics across all IRFs and the degree to which each IRF has an effect on functional change that differs from that of the average IRF.

5) Calculate an expected change in mobility score for each patient using a statistical model that estimates the average predictive effect of the patient characteristics across all IRFs on the basis of each patient's admission characteristics (risk adjustors)

6) Calculate an average predicted change in mobility score for each IRF. This is the IRF's predicted change in mobility score.

7) Calculate an average expected change in mobility score for each IRF. This is the IRF's expected change in mobility score.

8) For each IRF, divide the IRF's predicted change score by the IRF's expected change score to create a predicted to expected ratio, and then multiply each IRF's ratio by the national average change in mobility function.

2.2.6 Risk Adjustment

Patients treated in IRFs vary in terms of primary diagnosis (i.e., impairment group), demographic characteristics, and co-existing conditions. Patients may also have different expected improvement in function on the basis of these factors. Therefore, this outcome measure is risk adjusted. Risk adjustment controls for specific patient characteristics (e.g., age or diagnosis) that may affect patients' outcomes when comparing facilities.

An initial, extensive set of risk adjustment variables was selected for this quality measure on the basis of a review of literature and empirical findings from the PAC PRD analyses[7] as well as input from the TEP convened by RTI.[10] Using this initial set of risk adjustment variables, we have been conducting regression analyses using PAC PRD data to help identify the best set of risk adjustors on the basis of regression coefficients, statistical significance, sample sizes, and other indicators. We also requested input on suggested risk adjustors as part of the public comment process, and are incorporating suggestions we received as we refine our risk adjustment models. Data on reliability of CARE variables used for risk adjustment can be found in the report titled *The Development and Testing of the Continuity Assessment Record and Evaluation (CARE) Item Set: Final Report on Reliability Testing: Volume 2 of 3*.[8] The risk adjustment modeling approach follows the guidelines described by Ash et al. in the paper titled "Statistical Issues In Assessing Hospital Performance."[14]

The current list of risk adjustment variables is outlined below. This list will be updated, as appropriate, on the basis of further analyses.

- **Age group at IRF admission**
  - Younger than 35 years
  - 35 to 44 years

  - 45 to 54 years
  - 55 to 64 years
  - 65 to 74 years
  - 75 to 84 years
  - 85 years of age or older (reference category)
- **Primary rehabilitation diagnosis**
  - Stroke (reference category)
  - Brain dysfunction
  - Spinal cord dysfunction
  - Other neurological conditions
  - Factures and other multiple trauma
  - Hip and knee replacement
  - Other orthopedic conditions (amputation, arthritis, hip fracture)
  - Cardiac, pulmonary conditions, and debility
- **Medical vs. surgical acute-care diagnosis**
- **Functional status before current illness/injury: indoor ambulation**
  - Dependent or some help
  - Independent or unknown (reference category)
- **Functional status before current illness/injury: stairs**
  - Dependent or some help
  - Independent, or unknown (reference category)
- **Wheelchair use before current illness/injury**
  - Yes
  - No, or unknown (reference category)
- **History of falls in the last year**
  - Yes
  - No, or unknown (category)
- **Presence of severe pressure ulcer at admission** (Stage 3, Stage 4, or Unstageable pressure ulcer)
- **Cognitive abilities: Brief Interview for Mental Status (BIMS) score**
  - Severely impaired
  - Moderately impaired

- Intact or borderline (reference category)

- **Bladder incontinence**
  - Always incontinent
  - Less than daily or daily, Continent or stress incontinence only or no urine output (reference category)

- **Communication: Understanding verbal content and Expression of ideas and wants**
  - Moderate to severe communication limitations: Rarely/never understands; or sometimes understands; or rarely/never expresses self; or speech is very difficult to understand; or frequently expresses difficulty
  - Mild to no communication limitations: Usually understands; or some difficulty with expression
  - Mild to no communication limitations: Understands; or expression without difficulty; or unable to assess; or unknown (reference category).

- **Baseline mobility function score and baseline mobility function score squared**

- **Time from illness/injury onset to rehabilitation admission**

- **Comorbidities** (hierarchical condition categories) e.g., chronic kidney disease and dialysis; septicemia and other severe infections; lung and other severe cancers; lymphoma; diabetes; delirium and encephalopathy; tetraplegia; paraplegia

2.3 Quality Measure: IRF Functional Outcome Measure: Discharge Self-Care Score for Medical Rehabilitation Patients

2.3.1 Summary Description

This quality measure calculates the percentage of patients who meet or exceed an expected discharge *self-care* score in IRFs.

2.3.2 Purpose/Rationale for Quality Measure

Given that the primary goal of rehabilitation is improvement in functional status, IRF clinicians have traditionally assessed and documented patients' functional status at admission and at discharge to evaluate the effectiveness of the rehabilitation care provided to individual patients, as well as the effectiveness of the rehabilitation unit or hospital overall.

Studies have shown differences in IRF patients' functional outcomes by geographic region, insurance type, and race/ethnicity after adjusting for key patient demographic characteristics and admission clinical status, which supports the need to monitor IRF patients' functional outcomes.

As stated prior, the quality measures described in this document focus on self-care and mobility activities. We recognize that inpatient rehabilitation programs focus on recovery across many areas of function at the level of body structure and function, activities, and participation. Additional research is needed to develop quality measures for other areas of functioning.

2.3.3 Population

Inclusion Criteria

The population for this measure includes IRF patients who are 21 years old or older.

Exclusion Criteria

Two exclusion criteria apply to the discharge in self-care function quality measure:

1) Patients with incomplete stays: It can be challenging to gather accurate discharge functional status data for patients who experience incomplete stays. Patients with incomplete stays include patients who are unexpectedly discharged to an acute-care setting (IPPS, CAH, IPF, or LTCH) because of a medical emergency; patients who die or leave an IRF against medical advice; and patients with a length of stay less than 3 days.

2) Patients in coma, persistent vegetative state, complete teraplegia, and locked-in syndrome are excluded, because they may have limited or less predictable self-care improvement.

3) Patients younger than age 21.

2.3.4 Functional Status Items

One important consideration about measuring functional status is that certain functional status activities may not be safe for the patient; relevant; or feasible to collect for all patients in all types of settings. For example, an IRF patient may initially have difficulty swallowing and be at risk for aspiration; therefore, the clinician may indicate that a functional activity did not occur for patients on the basis of their clinical status because it is not safe or feasible for the patient to perform the activity.

In a recent TEP meeting convened by RTI, participants were asked to describe functional activities assessed by clinicians in their own facilities, as well as functional activities assessed by clinicians working in rehabilitation facilities that apply best practices. Participants noted that although some of the more challenging activities are not assessed at admission, it is important to assess them at discharge, especially for patients returning to a community-based setting.

The following functional activities are assessed and rated at the time of admission and discharge:

Self-Care Items

Eating: The ability to use suitable utensils to bring food to the mouth and swallow food once the meal is presented on a table/tray. Includes modified food consistency.

Oral hygiene: The ability to use suitable items to clean teeth.

Toilet hygiene: The ability to maintain perineal hygiene; ability to adjust clothes before and after using toilet, commode, bedpan, or urinal.

Shower/bathe self: The ability to bathe self in shower or tub, including washing, rinsing, and drying self. Does not include transferring in or out of tub/shower.

Upper body dressing: The ability to put on and remove shirt or pajama top. Includes buttoning, if applicable.

Lower body dressing: The ability to dress and undress below the waist, including fasteners. Does not include footwear.

Putting on/taking off footwear: The ability to put on and take off socks and shoes or other footwear that are appropriate for safe mobility.

Self-Care Rating Scale: Codes and Code Definitions

6. **Independent** – Patient completes the activity by himself/herself with no assistance from a helper.
5. **Setup or clean-up assistance** – Helper SETS UP or CLEANS UP; patient completes activity. Helper assists only prior to or following the activity.
4. **Supervision or touching assistance** – Helper provides VERBAL CUES or TOUCHING/ STEADYING assistance as patient completes activity. Assistance may be provided throughout the activity or intermittently.
3. **Partial/moderate assistance** – Helper does LESS THAN HALF the effort. Helper lifts, holds, or supports patient's trunk or limbs, but provides less than half the effort.

2. **Substantial/maximal assistance** – Helper does MORE THAN HALF the effort. Helper lifts or holds patient's trunk or limbs and provides more than half the effort.

1. **Dependent** – Helper does ALL of the effort. Patient does none of the effort to complete the task. Or, the assistance of 2 or more helpers is required for the patient to complete the activity.

If activity was not attempted code:

88. Not attempted due to medical condition or safety concerns

09. Not applicable

07. Patient refused

99. Not a patient goal (use at discharge only)

2.3.5 Quality Measure Calculation

The following steps are used to calculate the measure:

1) Sum the scores of the discharge self-care items to create a discharge self-care score for each patient, after 'activity did not occur' values are recoded. (range: 7 to 42). This is the patient's observed discharge score.

2) Calculate an expected discharge self-care score for each IRF patient using a statistical model that estimates both the average predictive effect of the patient demographic and admission clinical characteristics across all IRFs.

3) Compare each patient's observed and expected discharge self-care score and classify the difference as
   a) observed discharge score is higher than the expected discharge score, or
   b) observed discharge score is the same as the expected discharge score, or
   c) observed discharge score is lower than the expected discharge score.

4) Sum the number of patient with observed discharge scores that are the same as or higher than the expected discharge score.

5) The numerator is the number of patients in an IRF with a discharge score that is the same as or higher than the national discharge score (calculated in Step 4).

6) The denominator is the total number of patients in the IRF.

7) The percent is calculated as the numerator divided by the denominator.

2.3.6 Risk Adjustment

Patients treated in IRFs vary in terms of primary diagnosis (i.e., impairment group), demographic characteristics, and co-existing conditions. Patients may also have different expected improvement in function on the basis of these factors. Therefore, this outcome measure is risk adjusted. Risk adjustment controls for specific patient characteristics (e.g., age or diagnosis) that may affect patients' outcomes when comparing facilities.

An initial, extensive set of risk adjustment variables was selected for this quality measure on the basis of a review of literature and empirical findings from the PAC PRD analyses[7] as well as input from the TEP convened by RTI.[10] Using this initial set of risk adjustment variables, we have been conducting regression analyses using PAC PRD data to help identify the best set of risk adjustors on the basis of regression coefficients, statistical significance, sample sizes, and other indicators. We also requested input on suggested risk adjustors as part of the public comment process, and are incorporating suggestions we received as we refine our risk adjustment models. Data on reliability of CARE variables used for risk adjustment can be found in the report titled *The Development and Testing of the Continuity Assessment Record and Evaluation (CARE) Item Set: Final Report on Reliability Testing: Volume 2 of 3*.[8]

The current list of risk adjustment variables is outlined below. This list will be updated, as appropriate, on the basis of further analyses.

- **Age group at IRF admission**
  - Younger than 35 years
  - 35 to 44 years
  - 45 to 54 years
  - 55 to 64 years
  - 65 to 74 years
  - 75 to 84 years
  - 85 years of age or older (reference category)
- **Primary Rehabilitation diagnosis**
  - Stroke (reference category)
  - Brain dysfunction,
  - Spinal cord dysfunction
  - Other neurological conditions
  - Factures and other multiple trauma
  - Hip and knee replacement
  - Other orthopedic conditions (amputation, arthritis, hip fracture)
  - Cardiac, pulmonary conditions, and debility
- **Medical vs. surgical acute-care diagnosis**
- **Functional status before current illness/injury: self-care**
  - Dependent
  - Some help
  - Independent, or unknown (reference category)

- **Functional status before current illness/injury: indoor ambulation**
  - Dependent or some help
  - Independent, or unknown (reference category)
- **Wheelchair use before current illness/injury**
  - Yes
  - No, or unknown (reference category)
- **Presence of severe pressure ulcer at admission** (Stage 3, Stage 4, or Unstageable pressure ulcer)
- **Cognitive abilities: Brief Interview for Mental Status (BIMS) score**
  - Moderately impaired
  - Severely impaired
  - Intact or borderline (reference category)
- **Bladder incontinence**
  - Always incontinent
  - Less than daily or daily incontinence
  - Continent or stress incontinence only or no urine output (reference category)
- **Communication: Understanding verbal content *and* Expression of ideas and wants**
  - Moderate to severe communication limitations: Rarely/never understands; or sometimes understands; or rarely/never expresses self; or speech is very difficult to understand; or frequently exhibits difficulty with expression
  - Mild to no communication limitations: Usually understands or understands; or some difficulty with expression; or expression without difficulty; or unable to assess; or unknown (reference category)
- **Baseline self-care function score and baseline self-care function score squared**
- **Time from illness/injury onset to rehabilitation admission**
- **Comorbidities** (hierarchical condition categories) e.g., chronic kidney disease and dialysis; metastatic cancer and acute leukemia; diabetes; delirium and encephalopathy; paraplegia; multiple sclerosis; Parkinson's and Huntington's disease; chronic ischemic heart disease.

2.4 Quality Measure: IRF Functional Outcome Measure: Discharge Mobility Score for Medical Rehabilitation Patients

2.4.1 Summary Description

This measure calculates the percentage of patients who meet or exceed an expected discharge *mobility* score.

2.4.2 Purpose/Rationale for Quality Measure

Given that the primary goal of rehabilitation is improvement in function, IRF clinicians have traditionally assessed and documented patients' functional status at admission and at discharge to evaluate the effectiveness of the rehabilitation care provided to individual patients, as well as the effectiveness of the rehabilitation unit or hospital overall.

Studies have shown differences in IRF patients' functional outcomes by geographic region, insurance type, and race/ethnicity after adjusting for key patient demographic characteristics and admission clinical status, which supports the need to monitor IRF patients' functional outcomes.

As stated prior, the quality measures described in this document focus on self-care and mobility activities. We recognize that inpatient rehabilitation programs focus on recovery across many areas of function at the level of body structure and function, activities, and participation. Additional research is needed to develop quality measures for other areas of functioning.

2.4.3 Population

Inclusion Criteria

The population for this measure includes IRF patients who are 21 years old or older.

Exclusion Criteria

Two exclusion criteria apply to the discharge mobility function score measure:

1) Patients with incomplete stays: It can be challenging to gather accurate discharge functional status data for patients who experience incomplete stays. Patients with incomplete stays include patients who are unexpectedly discharged to an acute-care setting (IPPS, CAH, or LTCH) because of a medical emergency; patients who die or leave an IRF against medical advice; and patients with a length of stay less than 3 days.

2) Patients younger than age 21.

2.4.4 Functional Status Items

One important consideration when measuring functional status is that certain functional status activities may not be safe, relevant, or feasible to collect for all patients in all types of settings. For example, an IRF patient may be unable to walk up and down stairs at the time of admission. Therefore, clinicians may indicate that a functional activity did not occur for patients

on the basis of their clinical status because it is not safe or feasible for the patient to perform the activity.

In a recent TEP meeting convened by RTI, participants were asked to describe functional activities assessed by clinicians in their own facilities, as well as functional activities assessed by clinicians working in rehabilitation facilities that apply best practices. Participants noted that although some of the more challenging activities are not assessed at admission, it is important to assess them at discharge, especially for patients returning to a community-based setting.

For the quality measure, the following functional activities are assessed and rated at the time of admission and discharge:

Mobility Items

Roll left and right: The ability to roll from lying on back to left and right side, and roll back to back.

Sit to lying: The ability to move from sitting on side of bed to lying flat on the bed.

Lying to sitting on side of bed: The ability to safely move from lying on the back to sitting on the side of the bed with feet flat on the floor, no back support.

Sit to stand: The ability to safely come to a standing position from sitting in a chair or on the side of the bed.

Chair/bed-to-chair transfer: The ability to safely transfer to and from a chair (or wheelchair).

Toilet transfer: The ability to safely get on and off a toilet or commode.

Car transfer: The ability to transfer in and out of a car or van on the passenger side. Does not include the ability to open/close door or fasten seat belt.

For patients who are walking, complete the following items:

Walk 10 feet: Once standing, the ability to walk at least 10 feet (3 meters) in room, corridor, or similar space.

Walk 50 feet with two turns: The ability to walk 50 feet and make two turns.

Walk 150 feet (45 m): Once standing, the ability to walk at least150 feet (45 meters) in corridor or similar space.

Walking 10 feet on uneven surfaces: The ability to walk 10 feet on uneven or sloping surfaces, such as grass or gravel.

1 step (curb): The ability to step over a curb or up and down one step.

4 steps: The ability to go up and down four steps, with or without a rail.

12 steps: The ability to go up and down 12 steps, with or without a rail.

Picking up object: The ability to bend/stoop from a standing position to pick up a small object, such as a spoon from the floor.

Mobility Rating Scale: Codes and Code Definitions

6. **Independent** – Patient completes the activity by himself/herself with no assistance from a helper.
5. **Setup or clean-up assistance** – Helper SETS UP or CLEANS UP; patient completes activity. Helper assists only prior to or following the activity.
4. **Supervision or touching assistance** –Helper provides VERBAL CUES or TOUCHING/ STEADYING assistance as patient completes activity. Assistance may be provided throughout the activity or intermittently.
3. **Partial/moderate assistance** – Helper does LESS THAN HALF the effort. Helper lifts, holds, or supports patient's trunk or limbs, but provides LESS THAN HALF the effort.
2. **Substantial/maximal assistance** – Helper does MORE THAN HALF the effort. Helper lifts or holds patient's trunk or limbs and provides MORE THAN HALF the effort.
1. **Dependent** – Helper does ALL of the effort. Patient does none of the effort to complete the activity. Or, the assistance of 2 or more helpers is required for the patient to complete the activity.

If activity was not attempted, code one of the following:

88. Not attempted due to medical condition or safety concerns

09. Not applicable

07. Patient refused

99. Not a patient goal (use at discharge only)

2.4.5 Measure Calculation

The following steps are used to calculate the measure:

1) Sum the scores of the discharge mobility items to create a discharge mobility score for each patient, after 'activity did not occur' values are recoded. (range: 15 to 90). This is a patient's observed discharge score.

2) Calculate an expected discharge mobility score for each IRF patient using a statistical model that estimates both the average predictive effect of the patient demographic and admission clinical characteristics across all IRFs.

3) Compare each patient's observed and expected scores and classify the difference as

   a) observed discharge score is higher than the expected discharge score, or
   b) observed discharge score is the same as the expected discharge score, or
   c) observed discharge score is lower than the expected discharge score.

4) Sum the number of patient records with observed discharge scores that are the same as or higher than the expected discharge scores.

5) The numerator is the number of patients in an IRF with an observed discharge score that is the same as or higher than the expected discharge score (calculated in Step 4).

6) The denominator is the total number of patients in the IRF.

7) The percent is calculated as the numerator divided by the denominator.

2.4.6 Risk Adjustment

Patients treated in IRFs vary in terms of primary diagnosis (i.e., impairment group), demographic characteristics, and co-existing conditions. Patients may also have different expected improvement in function on the basis of these factors. Therefore, this outcome measure is risk adjusted. Risk adjustment controls for specific patient characteristics (e.g., age or diagnosis) that may affect patients' outcomes when comparing facilities.

An initial, extensive set of risk adjustment variables was selected for this quality measure on the basis of a review of literature review, empirical findings from the PAC PRD analyses[7] and input from the TEP convened by RTI.[9] Using this initial set of risk adjustment variables, we have been conducting regression, and other analyses using PAC PRD data to help identify the best set of risk adjustors on the basis of regression coefficients, statistical significance, sample sizes, and other indicators. We also requested input on suggested risk adjustors as part of the public comment process, and are incorporating suggestions as we refine our risk adjustment models. Data on reliability of CARE variables used for risk adjustment can be found in the report titled *The Development and Testing of the Continuity Assessment Record and Evaluation (CARE) Item Set: Final Report on Reliability Testing: Volume 2 of 3*. Risk adjustment modeling is currently in progress; we will examine other statistical approaches, such as hierarchical linear modeling to develop the final risk adjustment model. Our risk adjustment approach follows White Paper guidelines on risk adjustment by the Committee of Presidents of Statistical Societies and the Centers for Medicare & Medicaid Services (CMS).

The current list of risk adjustment variables is outlined below. This list will be updated, as appropriate, on the basis of further analyses.

- **Age group at IRF admission**
  - Younger than 35 years
  - 35 to 44 years
  - 45 to 54 years
  - 55 to 64 years
  - 65 to 74 years
  - 75 to 84 years
  - 85 years of age or older (reference category)

- **Primary rehabilitation diagnosis**
  - Stroke (reference category)
  - Brain dysfunction,
  - Spinal cord dysfunction
  - Other neurological conditions
  - Factures and other multiple trauma
  - Hip and knee replacement
  - Other orthopedic conditions (amputation, arthritis, hip fracture)
  - Cardiac, pulmonary conditions and debility
- **Medical vs. surgical acute-care diagnosis**
- **Functional status before current illness/injury: indoor ambulation**
  - Dependent or some help
  - Independent, or unknown (reference category)
- **Functional status before current illness/injury: stairs**
  - Dependent or some help
  - Independent, or unknown (reference category)
- **Wheelchair use before current illness/injury**
  - Yes
  - No or unknown (reference category)
- **History of falls in the last year**
  - Yes
  - No or unknown (reference category)
- **Presence of severe pressure ulcer at admission** (Stage 3, stage 4 or Unstageable pressure ulcer)
- **Cognitive abilities: BIMS score**
  - Severely impaired
  - Moderately impaired
  - Intact or borderline (reference category)
- **Bladder incontinence**
  - Always incontinent
  - Less than daily or daily, continent or stress incontinence only or no urine output (reference category)

- **Communication: Understanding verbal content and Expression of ideas and wants**
  - Moderate to severe communication limitations: Rarely/never understands; or sometimes understands; or rarely/never expresses self; or speech is very difficult to understand; or frequently exhibits difficulty with expression
  - Mild communication limitations: Usually understands; or some difficulty with expression
  - No communication limitations: Understands; or expression without difficulty (reference category)
- **Baseline mobility function score and baseline mobility function score squared**
- **Time from illness/injury onset to rehabilitation admission**
- **Comorbidities** (hierarchical clinical conditions)—for example, chronic kidney disease and dialysis; septicemia and other severe infections; lung and other severe cancers; lymphoma; diabetes; delirium and encephalopathy; tetraplegia; paraplegia

REFERENCES

1. World Health Organization. *International Classification of Functioning, Disability and Health*: Retrieved from http://apps.who.int/classifications/icfbrowser/ 2001.

2. Gage B, Constantine R, Aggarwal J, et al. *The Development and Testing of the Continuity Assessment Record and Evaluation (CARE) Item Set: Final Report on the Development of the CARE Item Set Volume 1 of 3*: RTI International;2012.

3. Gage B, Deutsch A, Smith L, et al. *The Development and Testing of the Continuity Assessment Record and Evaluation (CARE) Item Set: Final Report on CARE Item Set and Current Assessment Comparisons Volume 3 of 3*: RTI International;2012.

4. Gage B, Morley M, Smith L, et al. *Post-Acute Care Payment Reform Demonstration: Final Report Volume1of 4*: RTI International;2012.

5. Gage B, Morley M, Smith L, et al. *Post-Acute Care Payment Reform Demonstration: Final Report Volume 2 of 4*: RTI International;2012.

6. Gage B, Morley M, Smith L, et al. *Post-Acute Care Payment Reform Demonstration: Final Report Volume 3 of 4*: RTI International;2012.

7. Gage B, Morley M, Smith L, et al. *Post-Acute Care Payment Reform Demonstration: Final Report Volume 4 of 4*: RTI International;2012.

8. Gage B, Smith L, Ross J, et al. *The Development and Testing of the Continuity Assessment Record and Evaluation (CARE) Item Set: Final Report on Reliability Testing Volume 2 of 3*: RTI International;2012.

9. Deutsch A, Kline T, Kelleher C, Lines LM, Laurie Coots M, Garfinkel D. *Analysis of Crosscutting Medicare Functional Status Quality Metrics Using the Continuity and Assessment Record and Evaluation (CARE) Item Set: Final Report*: RTI International;2012.

10. RTI International. *Summary of Feedback from the Technical Expert Panel on the Development of Cross - Setting Functional Status Quality Measures*2014.

11. National Committee on Vital and Health Statistics Subcommittee on Health. *Classifying and Reporting Functional Status*2001.

12. Reistetter TA, Karmarkar AM, Graham JE, et al. Regional variation in stroke rehabilitation outcomes. *Arch Phys Med Rehabil.* Jan 2014;95(1):29-38.

13. O'Brien SR, Xue Y, Ingersoll G, Kelly A. Shorter length of stay is associated with worse functional outcomes for Medicare beneficiaries with stroke. *Physical therapy.* Dec 2013;93(12):1592-1602.

14. Ash AS, Fienberg SE, Louis TA, Normand S-LT, Stukel TA, Utts J. Statistical Issues in Assessing Hospital Performance. 2012.

APPENDIX A:
RELIABILITY AND VALIDITY TESTING

A.1 Overview of Reliability and Validity Testing

The functional assessment items used in the four inpatient rehabilitation facility (IRF) functional status quality measures are from the Continuity Assessment Record and Evaluation (CARE) Item Set. The CARE Item Set was designed to standardize assessment of patients' status across acute and post-acute settings, including IRFs, long-term care hospitals (LTCHs), skilled nursing facilities (SNFs), and home health agencies (HHAs). The functional status items on the CARE Item Set are daily activities that clinicians assess at the time of admission and/or at discharge to determine patients' needs, evaluate progress, and prepare for a transition home or another setting.

The goal of reliability testing is to ensure that items on an assessment obtain consistent results when administered or used by different clinicians. Validity testing examines whether an item or scale measures what it is intended to measure. The CARE functional status items underwent reliability testing at the item- and scale-level in multiple types of providers in conjunction with the Post-Acute Care Payment Reform Demonstration (PAC PRD). Item-level testing included inter-rater reliability testing within facilities and the use of videotaped standardized patients for inter-rater reliability testing across facilities/care settings. Additional testing focused on the items and scales and included internal consistency, factor analysis, and Rasch analysis. A brief summary of this testing is provided below; full reports describing the testing are available at http://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/Post-Acute-Care-Quality-Initiatives/CARE-Item-Set-and-B-CARE.html.

A.2 Traditional Inter-rater Reliability Study

The reliability of the functional items was tested in a subset of 34 providers from each of the five levels of care (acute hospitals, HHAs, IRFs, LTCHs, and SNFs) distributed across 11 geographic areas. Each provider completed a duplicate CARE Item Set (admission or discharge assessment) on 15–20 patients included in the PAC PRD (10–15 patients in the home health setting), in accordance with the guidelines and protocols.

Providers were asked to enroll a convenience sample of a set number of Medicare patients each month, representing a range of function and acuity. The overall patient sample size for each of the functional items was 450 for self-care items and 449 for mobility items (448 for transfers). After exclusions for missing data (unknown/not attempted/inapplicable), the effective sample sizes for the reliability testing were as follows:

- Eating: 401
- Oral hygiene: 414
- Toilet hygiene: 416
- Upper body dressing: 420
- Lower body dressing: 413
- Lying to sitting on the side of the bed: 412

- Sitting to standing: 387
- Chair/bed to chair transfer: 392
- Toilet transfer: 361
- Walk 150 feet: 68
- Walk once standing: 52
- Wheel in room: 46

The inter-rater reliability study included patients who were assessed by two different clinicians (raters), and the agreement of the clinicians' rating was calculated. Clinicians were instructed to have pairs of raters complete both patient assessments at the same time. Responses to items were obtained by direct observation of the patient by the clinician, and occasionally, supplemented by one or more of the following predetermined, matched methods: patient interviews (with each team member taking turns conducting and observing patient interviews); interviews with relatives/caregivers of the patient for certain items; and/or interviews with staff caring for the patient and/or chart review. Rater pairs were instructed to determine in advance which methods would be used to score the particular CARE items and to have both raters use the same methods. Raters were encouraged to divide hands-on assistance to the patient as evenly as possible for items that required hands-on assistance. Raters were instructed not to discuss item scoring during the assessment, nor to share item scores until the data were entered into the study database and finalized. Providers submitted data via the online CARE application for both assessments in each pair.

For categorical items, kappa statistics (kappa) indicate the level of agreement between raters using ordinal data, taking into account the role of chance agreement. The ranges commonly used to judge reliability based on kappa are as follows: ≤ 0 = poor; 0.01–0.20 = slight; 0.21–0.40 = fair; 0.41–0.60 = moderate; 0.61–0.80 = substantial; and 0.81–1.00 = almost perfect.

For categorical items with only two responses available, RTI International calculated only unweighted kappas. For items with more than two responses, RTI calculated both weighted and unweighted kappas. Unweighted kappa assumes the same "distance" between every one-unit difference in response across an ordinal scale. RTI used Fleiss-Cohen weights, or quadratic weights, which approximate the intra-class correlation coefficient and are commonly used for calculating weighted kappas. This choice of weighting is consistent with prior analyses of assessment reliability, where the method for developing weights was specified.[1,2] Fleiss-Cohen weights put lower emphasis on disagreements between responses that fall near each other on an item scale. It should also be noted that the value of kappa can be influenced by the prevalence of the outcome or characteristic being measured. If the outcome or characteristic is rare, the kappa will be low because kappa attributes the majority of agreement among raters to chance. Kappa is also influenced by bias, and if the effective sample size is small, variation may play a role in the results. Hence, we report both weighted and unweighted kappas to give the range of agreement found under the two sets of assumptions.

Additionally, RTI calculated a separate set of kappa statistics (unweighted and weighted, where applicable) for items where additional responses outside of an ordinal scale were available (letter codes) and were set to missing.

For the traditional reliability study, kappa statistics indicated substantial agreement among raters. The weighted kappa values for the self-care items range between 0.798 for eating to 0.869 for upper-body dressing. Unweighted kappas ranged from 0.598 for oral hygiene to 0.634 for upper-body dressing. Provider-specific analyses of core self-care items show similar agreement to the overall estimates. The lower-body dressing item had the highest overall weighted kappa (0.855), whereas the eating item had the lowest (0.798). Unweighted overall kappas ranged from 0.636 (toileting) to 0.598 (oral hygiene). Acute hospitals had the highest weighted kappas across all self-care items.

The weighted kappa values for the mobility items ranged between 0.558 for walk 150 feet to 0.901 for sitting to standing and chair/bed to chair transfer. Unweighted kappas ranged from 0.667 for walk once standing to 0.762 for sit to stand. Provider-specific analyses of core mobility items show similar agreement to the overall estimates. The sit-to-stand and chair transfer items both had a weighted kappa of 0.901, whereas the lying to sitting item had a weighted kappa of 0.855. Unweighted overall kappas ranged from 0.693 (lying to sitting) to 0.762 (sitting to standing).

A.3 Videotaped Standardized Patients Reliability Study

For the video reliability study, which was designed to examine the level of clinician agreement across care settings, clinicians in each setting were asked to assess "standardized" patients presented through a videotape of a patient assessment. This ensured that the same information was presented to each clinician and allowed examination of differences in scoring effects among different clinicians examining the "same" patient.

The patient "case studies" in each of the videos varied in terms of medical complexity, functional abilities, and cognitive impairments. The nine videos included patients classified as high, medium, or low ability/complexity for each of these three areas. Each facility or agency received three videos, one of which demonstrated one of the following elements: cognitive impairments, skin integrity problems, a wheelchair-dependent patient, and a variety of mid-level functional activities. The mid-level functional activities were considered to be the most challenging for clinicians to score and are thus of particular interest in establishing reliability. Each clinician involved in the video study watched three videos and assessed the patients according to the study guidelines and protocols. Each video was approximately 20 minutes long and had a corresponding item set arranged in the sequence in which the items appeared in the video.

The sample included 28 providers (550 assessments), which included 3 acute hospitals (15 assessments [3%]); 9 HHAs (118 assessments [22%]); 8 IRFs (237 assessments [43%]); 3 LTCHs (114 assessments [21%]); and 5 SNFs (66 assessments [12%]). Participating providers included case managers (6% of assessments), occupational therapists (14% of assessments), physical therapists (21% of assessments), registered nurses (47% of assessments), speech therapists (5% of assessments), and others, mostly licensed practical nurses (LPNs; 8% of assessments).

Two main analytic approaches were used for assessing the video reliability of the CARE items, adhering closely to the methods used by Fricke et al. in their video reliability study of the

FIM® instrument.[3] First, percent agreement with the mode response was calculated for each CARE item included in at least one of the nine videos. Unlike the approach used by Fricke et al., RTI did not consider agreement at one response level above and below the mode, and instead used a stricter approach looking at direct modal agreement only. In the second approach, percent agreement with the internal clinical team's consensus response was also calculated. This second measure not only gives an indication of item reliability, but also reflects training consistency for the providers.

The video reliability study indicated substantial agreement with the mode and clinical team among all items, typically upwards of 70%. The notable exception to this trend exists among the clinicians in the "Other" category (mostly LPNs); they consistently had the lowest levels of agreement among all core self-care items, ranging from 50 to 72%. For the toileting and dressing items, the agreement with the clinical team was lower than with the mode. This occurred because the clinical team response differed from the mode for these three items in either one or two videos. Nonetheless, because the clinical team response and mode were identical on most of the videos, agreement was still quite high for these items. In general, study clinicians had responses on average that agreed with the expert clinical team or were slightly lower.

The video reliability study indicated substantial agreement with the mode and clinical team for the Lying-to-Sitting, Sit-to-Stand, Chair/Bed to Chair Transfer, and Toilet Transfer items (greater than 76%). Although rates of agreement with the mode and clinical team response were generally identical, for the Toilet Transfer item, the clinical team agreement is slightly lower. The items for walking and wheeling distances showed more variable levels of agreement across disciplines, with overall agreement generally in the moderate range (50–78%). For the Walk In Room item, there was a notable decrease in the agreement with the clinical team compared to agreement with the mode. This occurred because in two of the four videos where this item was assessed, the clinical team response differed from the mode.

A.4 Scale-level Reliability Results: Internal Consistency

In addition to item-level reliability testing, we examined internal consistency, which provides a general assessment of how well the items interrelate within a domain or subscale. Internal consistency is assessed using the Cronbach's alpha coefficient, which is the average correlation of all possible half-scale divisions. Cronbach's alpha is a statistic frequently assessed when instrument or scale psychometrics are published. The Cronbach's alpha reliability estimate ranges from zero to one, with an estimate of zero indicating that there is no consistency of measurement among the items, and one indicating perfect consistency. Many cutoff criteria exist to determine whether or not a scale shows good consistency or whether the items "hang together" well. General consensus is that Cronbach's alpha should be at least 0.70 for an adequate scale for group-level decisions, and alphas closer to 1 indicate a good scale.[4]

Assessments of individual self-care and mobility subscales at both admission and discharge tend to show good reliability statistics (Cronbach's Alpha of at least 0.80) within their specified subscales. Reliability estimates by provider type show that the functional status items maintain a very high internal consistency. In addition, no one provider type appears to have

FIM® is a trademark of Uniform Data System for Medical Rehabilitation, a division of UB Foundation Activities, Inc.

reliability estimates higher or lower than the rest, indicating similarity of CARE usage with respect to internal consistency.

The following table shows the findings from the Cronbach's alpha internal consistency evaluation mentioned above.

Table A-1
CARE functional status internal consistency reliability summary by provider type

CARE analytic set	Overall alpha	HHA alpha	SNF alpha	IRF alpha	LTCH alpha
Self-Care	0.96	0.94	0.95	0.95	0.96
Mobility	0.96	0.94	0.95	0.96	0.97

A.5 Scale-level Reliability and Validity Testing: Rasch Analysis

Because we are measuring a latent trait—a concept that is not measured directly, but that relies on activities that can be directly observed—we used the one-parameter Rasch model to gain a better understanding of the functional status activities. More specifically, we examined the order of functional status items (from least challenging to most challenging) that characterize the concepts of the self-care and mobility.

Rasch analysis uses the scores from the functional assessment items to create the equivalent of a functional status "ruler" (i.e., scale). Rasch analysis uses the available data to estimate a person's location along the "ruler;" therefore, analyses can be conducted if some data are missing. Rasch analysis can also inform the optimal selection of key items in order to construct functional status scales that sufficiently span an entire range of patient functioning, so that both the least able and most able (lowest- and highest-functioning) patients are adequately measured. In addition, Rasch analysis can indicate where items overlap or are redundant in terms of the level of function they capture.

Rasch analysis has been used to examine the FIM® instrument,[5-8] the Minimum Data Set (MDS),[9] and the Outcome and Assessment Information Set (OASIS).[10] Rasch analysis has also been used to examine the extent to which existing functional assessment instruments (e.g., the FIM® instrument, MDS 2.0) capture the same construct.[11]

Rasch measurement is based on a probabilistic model that describes the association between a person's underlying ability level and probability of a particular item response, and summarizes a patient's position along a "ruler" that represents a latent trait or concept (e.g., self-care or mobility).[12] In essence, the Rasch analysis creates a ruler based on the domain measured (e.g., mobility) that can be used to assess the abilities of the patients. The analysis also provides information on the hierarchy of item difficulty (from easy to hard) that can be used to evaluate the construct validity of a set of items. In addition, the Rasch analysis provides information about the level of challenge associated with each item rating scale ("dependent" through "independent"). For example, an item with a low difficulty estimate (e.g., eating) would be more likely to be completed with little or no help by patients items that are more challenging (e.g., 12 step), where most patients would find completing this activity challenging. Finally, the Rasch

analysis can provide information on items that do not fit into the single theorized concept through "item misfit" statistics, which may indicate that the item needs further evaluation before being included on future administrations of the subscale. The infit mean square is an indicator of the degree to which patient responses are similar to what would be expected (i.e., predicted) by the measurement model. The acceptable range is generally 0.6 to 1.4. If the item values are above this range, it reflects that person response patterns are erratic, generally suggesting that the item is not measuring the same construct as other items. Infit mean squares above 1.4 are considered to be unacceptably unexpected[13] and indicate that the item most likely does not reflect the same construct as the other items included in the scale; for example, a need for assistance with self-care.

RTI used Rasch analysis to examine the extent to which the items worked together to define a coherent concept. This was conducted separately for the self-care and mobility items. Item fit statistics were examined as an indication of how well all items work together to describe the overall construct (self-care or mobility). The Rasch analysis provides insight into how the items work together as a subscale, including the hierarchy of item difficulty (ordering from easy to difficult) and item fit to the model.

Examinations of these Rasch analysis results reveal that the mobility and self-care item hierarchies make sense clinically and that the operational definitions of the constructs maintain general stability from admission to discharge. Some items have fit statistics outside the acceptable range (e.g., pick up object from floor), but the Technical Expert Panel members notes that this is an important assessment given the risk of falls.

RTI examined how well the items selected measure the persons in the data set for both self-care and mobility items. RTI examined the extent to which person response patterns fit the assumptions of the measurement model using the same range of infit statistics identified above. RTI examined the extent to which persons are effectively measured (ceiling and floor effects) in each setting overall and for admission and discharge time points. The mobility and self-care items were found to be well targeted to the range of patient ability sampled within this PAC population.

RTI established that the six steps of the CARE rating scale are operating as intended, both overall and for individual items on the self-care and mobility subscales. The probability that a person will be scored on a particular rating scale step varies depending on the functional ability of the person. That is, very able people will be more likely to be scored as "5" and "6" than as "1" and "2." Looking empirically at these distributions, we should see the transitions from one step to the next (called thresholds) proceed monotonically and distinctly across the range of person abilities. In other words, there should always be some point along the range at which each rating-scale step is more probable than another step. When a rating-scale step is not more probable at any point, it suggests that raters are not able to use that step to consistently distinguish patient ability at that level.

REFERENCES

1. Hirdes JP, Smith TF, Rabinowitz T, et al. The Resident Assessment Instrument-Mental Health (RAI-MH): inter-rater reliability and convergent validity. *J Behav Health Serv Res.* 29(4):419-432, 2002

2. Streiner DL, Norman GR. Health measurement scales: a practical guide to their development and use. *Oxford University Press*, 1995.

3. Fricke J, Unsworth C, Worrell D. Reliability of the Functional Independence Measure with Occupational Therapists. *Australian Occupational Therapy Journal* 40(1):7-15, 1993.

4. Aron A, Aron EN *Statistics for Psychology*. 2nd ed. Upper Saddle River, NJ: Prentice Hall, 1999.

5. Granger CV, Hamilton BB, Linacre JM, et al. Performance profiles of the functional independence measure. *Am J Phys Med Rehabil.* 72(2):84-89, 1993.

6. Linacre JM, Heinemann AW, Wright BD, et al. The structure and stability of the Functional Independence Measure. *Archives of Physical Medicine & Rehabilitation.*75(2):127-132, 1994.

7. Wright BD, Linacre JM, Smith RM, et al. FIM measurement properties and Rasch model details. *Scandinavian Journal of Rehabilitation Medicine*, 29(4):267-272, Dec. 1997.

8. Heinemann AW, Linacre JM, Wright BD, et al. Relationships between impairment and physical disability as measured by the functional independence measure. *Arch Phys Med Rehabil.* 74(6):566-573, 1993.

9. Wang YC, Byers KL, Velozo CA. Rasch analysis of Minimum Data Set mandated in skilled nursing facilities. *J Rehabil Res Dev.* 45(9):1385-1399, 2008.

10. Fortinsky RH, Garcia RI, Joseph Sheehan T, et al. Measuring disability in Medicare home care patients: application of Rasch modeling to the outcome and assessment information set. *Med Care.* 41(5):601-615, 2001.

11. Velozo CA, Byers KL, Wang YC, et al. Translating measures across the continuum of care: using Rasch analysis to create a crosswalk between the Functional Independence Measure and the Minimum Data Set. *J Rehabil Res Dev.* 44(3):467-478, 2007.

12. Wright BD, Stone MH. Best Test Design. *Rasch Measurement*. 1979.

13. Wright BD, Linacre JM, Gustafson J, et al. Reasonable mean-square fit values. *Rasch Measurement Transactions.* 8(3):370, 1994.