Discharge Function Score for Inpatient Rehabilitation Facilities (IRFs)

Technical Report

February 2023

Prepared for:
Center for Clinical Standards and Quality
Centers for Medicare & Medicaid Services
7500 Security Boulevard
Baltimore, MD 21244-1850
Contract No. 75FCMC18D0015, Task Order 75FCMC19F0003

Acumen, LLC
500 Airport Blvd., Suite 365
Burlingame, CA 94010

TABLE OF CONTENTS

LIST OF TABLES AND FIGURES

1 INTRODUCTION

The Patient Protection and Affordable Care Act (ACA) of 2010[1] and Improving Medicare Post-Acute Care Transformation Act (IMPACT) of 2014[2] require the Secretary to establish public reporting requirements for quality measures for inpatient rehabilitation facilities (IRFs) using standardized patient assessment data elements. As part of this mandate, the Centers for Medicare & Medicaid Services (CMS) has contracted with Acumen, LLC to develop a cross-setting functional outcome measure to be used in the IRF Quality Reporting Program (QRP) under the *Quality Measure & Assessment Instrument Development & Maintenance & QRP Support* contract (75FCMC18D0015/Task Order 75FCMC19F0003).

Measuring functional status of IRF patients can provide valuable information about an IRF's quality of care. Physical function predicts several outcomes including successful discharge to the community and re-hospitalization rates.[3,4,5] Several studies have reported that IRF care can improve patients' motor function at discharge for patients with various diagnoses, including traumatic brain injury and stroke.[6,7,8,9] Providers can intervene to improve patients' functional

[1] Section 3004(b) of the Patient Protection and Affordable Care Act of 2010, Pub.L. 111-148
[2] Amendment Section 1899B to the Social Security Act, Pub.L. 113-185
[3] Minor M, Jaywant A, Toglia J, Campo M, O'Dell MW. Discharge Rehabilitation Measures Predict Activity Limitations in Patients with Stroke Six Months after Inpatient Rehabilitation. Am J Phys Med Rehabil. 2021 Oct 20. doi: 10.1097/PHM.0000000000001908. Epub ahead of print. PMID: 34686630.
[4] Deutsch A, Palmer L, Vaughan M, Schwartz C, McMullen T. Inpatient Rehabilitation Facility Patients' Functional Abilities and Validity Evaluation of the Standardized Self-Care and Mobility Data Elements. Arch Phys Med Rehabil. 2022 Feb 11:S0003-9993(22)00205-2. doi: 10.1016/j.apmr.2022.01.147. Epub ahead of print. PMID: 35157893.
[5] Li CY, Haas A, Pritchard KT, Karmarkar A, Kuo YF, Hreha K, Ottenbacher KJ. Functional Status Across Post-Acute Settings is Associated With 30-Day and 90-Day Hospital Readmissions. J Am Med Dir Assoc. 2021 Dec;22(12):2447-2453.e5. doi: 10.1016/j.jamda.2021.07.039. Epub 2021 Aug 30. PMID: 34473961; PMCID: PMC8627458.
[6] Evans E, Krebill C, Gutman R, Resnik L, Zonfrillo MR, Lueckel SN, Zhang W, Kumar RG, Dams-O'Connor K, Thomas KS. Functional motor improvement during inpatient rehabilitation among older adults with traumatic brain injury. PM R. 2021 May 21:10.1002/pmrj.12644. doi: 10.1002/pmrj.12644. Epub ahead of print. PMID: 34018693; PMCID: PMC8606011.
[7] Kowalski RG, Hammond FM, Weintraub AH, et al. Recovery of Consciousness and Functional Outcome in Moderate and Severe Traumatic Brain Injury. JAMA Neurol. 2021;78(5):548-557. doi:10.1001/jamaneurol.2021.0084
[8] Li CY, Karmarkar A, Kuo YF, Haas A, Ottenbacher KJ. Impact of Self-Care and Mobility on One or More Post-Acute Care Transitions. J Aging Health. 2020;32(10):1325-1334. doi:10.1177/0898264320925259
[9] O'Dell MW, Jaywant A, Frantz M, Patel R, Kwong E, Wen K, Taub M, Campo M, Toglia J. Changes in the Activity Measure for Post-Acute Care Domains in Persons With Stroke During the First Year After Discharge From Inpatient Rehabilitation. Arch Phys Med Rehabil. 2021

outcomes by adopting a patient-centered care plan that accounts for each patient's unique circumstances.[10,11]

The Discharge Function Score measure determines how successful each IRF is at achieving an expected level of functional ability for its patients at discharge. This measure was placed on the 2022 Measures Under Consideration (MUC) List as the 'Cross-Setting Discharge Function Score', and CMS subsequently modified the measure's name to the 'Discharge Function Score.' An expectation for discharge function score is built for each IRF stay by accounting for patient characteristics that impact their functional status. The final Discharge Function Score for a given IRF is the proportion of that IRF's stays where a patient's observed discharge score meets or exceeds their expected discharge score. IRFs with low scores are not producing the functional gains that they could be for a larger share of their patients. The measure provides actionable feedback to IRFs that has the potential to hold providers accountable and encourage them to improve the quality of care they deliver. This measure also promotes patient wellness, encourages the provision of adequate therapy to help prevent adverse outcomes (e.g., re-hospitalization), and increases the transparency of quality of care in the IRF setting. The Discharge Function Score measure adds value to the IRF QRP function measure portfolio by using specifications that allow for better comparisons across post-acute care (PAC) settings, considering both self-care and mobility activities in the function score, and refining the approach to addressing missing item scores.

Input from a variety of stakeholders has been taken into consideration throughout the measure development process. Feedback was sought and considered from patients and caregivers on the salience of the measure concept and from Technical Expert Panels (TEPs) on the appropriate specifications for the cross-setting measure.

This report presents the technical measure specifications for the Discharge Function Score measure. Section 2 provides an overview of the measure and a high-level summary of the key features of the measure that are described in detail in the remaining sections of the document. Section 3 describes the methodology used to construct the Discharge Function Score measure including its data sources, study population, measure outcome, and steps for calculating the final measure score. Section 4 discusses Discharge Function Score measure testing, including the measure's reportability, variability, reliability, and validity testing results. Lastly, the

[10] Cogan AM, Weaver JA, McHarg M, Leland NE, Davidson L, Mallinson T. Association of Length of Stay, Recovery Rate, and Therapy Time per Day With Functional Outcomes After Hip Fracture Surgery. JAMA Netw Open. 2020 Jan 3;3(1):e1919672. doi: 10.1001/jamanetworkopen.2019.19672. PMID: 31977059; PMCID: PMC6991278.

[11] Evans E, Krebill C, Gutman R, Resnik L, Zonfrillo MR, Lueckel SN, Zhang W, Kumar RG, Dams-O'Connor K, Thomas KS. Functional motor improvement during inpatient rehabilitation among older adults with traumatic brain injury. PM R. 2021 May 21:10.1002/pmrj.12644. doi: 10.1002/pmrj.12644. Epub ahead of print. PMID: 34018693; PMCID: PMC8606011.

Appendix includes risk adjustment model results and supporting information for the statistical imputation models used to estimate missing item scores.

2 OVERVIEW

This section provides an overview of basic descriptive information on the Discharge Function Score measure, summarizing the key points contained in the rest of the document. A more detailed explanation of the measure specifications is available in Section 3.

2.1 Measure Name

Discharge Function Score

2.2 Measure Type

Outcome Measure

2.3 Care Setting

IRF

2.4 Data Source

Inpatient Rehabilitation Facility-Patient Assessment Instrument (IRF-PAI)

2.5 Brief Description of Measure

The Discharge Function Score calculates the percent of IRF patients who achieve a risk-adjusted expected function score at discharge. Functional status is measured through Section GG of IRF-PAI assessments, which evaluates a patient's capacity to perform daily activities related to self-care (GG0130) and mobility (GG0170). Coefficients from a risk adjustment model controlling for admission function score, age, and patient clinical characteristics are used to determine an expected discharge function score for each IRF stay. The provider score is calculated as the following proportion:

$$\frac{Number\ of\ provider's\ stays\ where\ observed\ \text{discharge score} \geq\ expected\ discharge\ score}{Total\ number\ of\ provider's\ stays} * 100$$

3 MEASURE SPECIFICATIONS

This section describes the methodology used to construct the Discharge Function Score. Section 3.1 describes the study window for the measure. Section 3.2 summarizes the data source used to calculate the measure score. Section 3.3 details the study population used for the measure denominator. Section 3.4 defines the discharge function outcome used for the measure numerator. Section 3.5 reviews the imputation methodology used to estimate missing item scores. Section 3.6 describes the risk adjustment model and variables used for risk adjustment. Section 3.7 presents the steps involved in calculating the final measure score.

3.1 Measure Time Period

This measure is calculated using 12 months (four quarters) of data. All IRF stays with a discharge date that falls within this target period, except those that meet the exclusion criteria (refer to Section 3.3.2 for details), are included in the measure.

3.2 Data Source

This measure uses data from the IRF-PAI. The IRF-PAI data are collected on all Medicare patients who receive services from an inpatient rehabilitation unit or hospital. This measure is calculated entirely using administrative data. There will be no additional data collection or submission burden for IRF providers as the data used in the measure are already collected on the IRF-PAI.

3.3 Denominator

The denominator is the total number of IRF stays with an IRF-PAI record in the measure target period, which do not meet the exclusion criteria.

3.3.1 Stay Construction

An IRF-PAI record is submitted when a patient is discharged from the IRF and includes both admission and discharge data. An IRF stay includes consecutive time in the facility starting with a patient's admission date through the patient's discharge date and is inclusive of interrupted stay days. An interrupted IRF stay is defined as those cases in which a Medicare beneficiary is discharged from the IRF and returns to the same IRF within three consecutive calendar days. The three consecutive calendar days begin with the day of the discharge from the IRF and end on midnight of the third day.

The target date for an IRF-PAI record is the discharge date. The target period for the measure is 12 months (4 quarters). To construct the IRF stays, all IRF stays with a target date within the target period are selected. IRF stays are sorted by the Provider Internal ID, Patient ID, Admission Date, Discharge Date, Correction Number, and IRF Assessment ID. For each unique

admission date, only the first record is selected to eliminate duplicates. If IRF stays for the same Provider Internal ID and Patient Internal ID are overlapping by more than one day, both stays are removed. If a patient has multiple eligible IRF stays with a discharge date within the target period, then each eligible stay is included in the measure.

3.3.2 Eligible Stays

The eligible stays for this measure are all IRF-PAI stays that do not meet the exclusion criteria during the target period. The IRF stay is excluded if any of the following are true:

- Patient has an incomplete stay. Patients with incomplete stays include patients who are discharged to an acute care setting (Short-stay Acute Hospital, Critical Access Hospital, Inpatient Psychiatric Facility, or Long-term Care Hospital) for longer than 3 calendar days; patients who die or leave an Inpatient Rehabilitation Facility (IRF) against medical advice; and patients with a length of stay less than three days.
  Rationale: When a patient has an incomplete stay, for example, the patient leaves urgently due to a medical emergency, it can be challenging to gather accurate discharge functional status data.
- Patient is in a coma, persistent vegetative state, or locked-in state, or has complete tetraplegia, severe anoxic brain damage, cerebral edema, or compression of brain.

  Rationale: These patients are excluded because they may have limited or less predictable mobility improvement with the selected items.
- Patient is younger than 18 years: Age in years is calculated based on the truncated difference between admission date and birth date, i.e., the difference is not rounded to the nearest whole number.

  Rationale: Patient under 18 years old are not included in the target population for this measure because pediatric IRF patients may have different patterns of care than adult patients.
- Patient is discharged to hospice (home or institutional facility).

  Rationale: Patient goals may change during the IRF stay, and functional improvement may no longer be a goal for a patient discharged to hospice.

3.4 Numerator

The numerator is the number of stays in an IRF with an observed discharge function score (Section 3.4.1) for Section GG function items that is equal to or higher than the calculated expected discharge function score (Section 3.4.2).

3.4.1 Observed Discharge Function Score

The observed discharge function score is the sum of individual function items at discharge. Section GG of each PAC assessment instrument other than Hospice includes standardized patient assessment data elements that measure functional status. The Discharge Function Score measure focuses on GG items that are currently available across these PAC settings (Table 1).

Table 1. Cross-Setting Function Item Set

Item	Item Description
GG0130A	Eating
GG0130B	Oral Hygiene
GG0130C	Toileting Hygiene
GG0170A	Roll Left and Right
GG0170C	Lying to Sitting on Side
GG0170D	Sit to Stand
GG0170E	Chair/Bed-to-Chair Transfer
GG0170F	Toilet Transfer
GG0170I	Walk 10 Feet
GG0170J	Walk 50 Feet with 2 Turns
GG0170R	Wheel 50 Feet with 2 Turns

Valid responses for GG items are reported in Table 2.

Table 2. GG Items Response

Category	GG Items Response	Response Description
Patient Functional Status Assessed	6	Independent
	5	Setup or clean-up assistance
	4	Supervision or touching assistance
	3	Partial/moderate assistance
	2	Substantial/maximal assistance
	1	Dependent
Activity Not Attempted (ANA) codes	7	Patient refused
	9	Not applicable
	10	Not attempted due to environmental limitations
	88	Not attempted due to medical condition or safety concerns
Other NA codes	^	Skip pattern
	-	Not assessed/no information

The following steps are used to determine the observed discharge function score for each stay:

Step 1: If the code for an item is between 1 and 6, then use code as the score for that item.

Step 2: If code for an item is 7, 9, 10, 88, dashed (-), skipped (^), or missing, then use statistical imputation to estimate the item score for that item (see Section 3.5).

Step 3: Sum scores across all items to calculate the total observed discharge function score. Different locomotion items are used if the patient uses a wheelchair than for the remaining patients.

Use 2 * Wheel 50 Feet with 2 Turns (GG0170R) score to calculate the total observed discharge function score for stays where (i) Walk 10 Feet (GG0170I) has an activity not attempted (ANA) code at both admission and discharge and (ii) either Wheel 50 Feet with 2 Turns (GG0170R) or Wheel 150 Feet (GG0170S) has a code between 1 and 6 at either admission or discharge. The remaining stays use Walk 10 Feet (GG0170I) + Walk 50 Feet with 2 Turns (GG0170J) to calculate the total observed discharge function score.

In either case, 10 items are used to calculate a patient's total observed discharge score and score values range from 10 – 60.

3.4.2 Expected Discharge Function Score

The expected discharge function score is determined by applying the regression equation determined from risk adjustment to each IRF stay. Risk adjustment controls for patient characteristics such as admission function score, age, and clinical conditions. Refer to Section 3.6 for details on risk adjustment.

3.5 Statistical Imputation

When an item score is missing because an ANA code, a dash (-), or a skip (^) has been recorded (henceforth referred to as NA) rather than a value of 1 to 6, item scores are estimated through statistical imputation. This approach refines the imputation method used for in-use IRF QRP Functional Outcome Measures: Change in Self-Care Score for Medical Rehabilitation Patients (NQF #2633) (CMS ID: I009.03), Discharge Self-Care Score for Medical Rehabilitation Patients (NQF #2635) (CMS ID: I011.03), Change in Mobility Score for Medical Rehabilitation Patients (NQF #2634) (CMS ID: I010.03), and Discharge Mobility Score for Medical Rehabilitation Patients (NQF #2636) (CMS ID: I012.03). The method used in these measures recodes all NAs to 1, which implicitly assumes all NA codes signify patients who are completely dependent on a functional activity. On average, patients who are coded as NA on a GG item at admission tend to score higher at discharge (if assessed) than patients who are coded as dependent on admission. Treating both types of patients the same in risk adjustment can lead to

less accurate expected discharge values for each of these types of patients. Statistical imputation allows NAs to take any value from 1 to 6, based on a patient's clinical characteristics and codes assigned on other GG items.

A separate statistical imputation model is constructed for each GG item used in the Discharge Function Score numerator (Section 3.4.1) at admission and discharge. Imputation models include the predictors used in risk adjustment (Section 3.6.2) and covariates for scores on other GG items (Step 3 below). Notably, imputation models use all GG items available in IRF to estimate missing scores for the subset of GG items used for the Discharge Function Score numerator (detailed imputation model results are available upon request). The following steps are used to generate imputed item scores for stays with NA codes. Note that these steps first describe imputing a single item at admission and then describe the relevant modifications for discharge and for the other items.

Step 1: Start with Eating (GG0130A). Identify eligible stays where the item score is not missing (i.e., had a score 1-6) at admission. These scores are used as the outcome (i.e., left-hand-side variable) of the admission imputation model for GG0130A.

Step 2: For each stay, determine whether to use walking or wheeling items in the imputation model.

a) If Walk 10 Feet (GG0170I) has an ANA code at both admission and discharge and either Wheel 50 Feet with 2 Turns (GG0170R) or Wheel 150 Feet (GG0170S) has a code between 1 and 6, then use wheeling items.

b) Otherwise, use walking items.

Step 3: Create variables for the imputation model reflecting how each item (g_2 through g_l) except Eating (GG0130A) was scored at admission. GG item scores are described as independent variables (i.e., on the right-hand side) by three variables, collectively referred to as g'. The first reflects a score of 1-6 when available (g), the second is an indicator variable taking a value of 1 if the item had an ANA code, dash, or missing value ($g*$), and the third is an indicator variable taking a value of 1 if the item was skipped ($g**$).

$$Function\ items: G \in \{g_2, \ldots, g_{10}\} \quad (1)$$

$$g' = [g, g^{*}, g^{**}] \tag{2}$$

$$g = \begin{cases} g, & g = \{1,2,3,4,5,6\} \\ 0, & otherwise \end{cases}$$

$$g^{*} = \begin{cases} 1, & g = \{7,9,10,88,\text{-}\} \\ 0, & otherwise \end{cases}$$

$$g^{**} = \begin{cases} 1, & g = \{\text{^}\} \\ 0, & otherwise \end{cases}$$

$$Function\ items\ with\ NA\ indicators : G' \in \{g'_2, \dots, g'_{10}\} \tag{3}$$

Step 4: Estimate an ordered probit model using the sample identified in Step 1.

Two types of predictors (i.e., right-hand-side variables) are used in the imputation method: clinical covariates (C) and function items with NA indicators (G') constructed in Step 3.

$$Clinical\ items := C \in \{c_1, \dots, c_k\} \tag{4}$$

$$Function\ items\ with\ NA\ indicators : G' \in \{g'_2, \dots, g'_{10}\} \tag{5}$$

The model we estimate for g_1, GG0130A, is

$$z_i = C_i\beta + G_i'\phi + \varepsilon_i \tag{6}$$

$$g_i = \begin{cases} 1, & z_i \le \alpha_1 \\ 2, & \alpha_1 < z_i \le \alpha_2 \\ 3, & \alpha_2 < z_i \le \alpha_3 \\ 4, & \alpha_3 < z_i \le \alpha_4 \\ 5, & \alpha_4 < z_i \le \alpha_5 \\ 6, & z_i > \alpha_5 \end{cases} \tag{7}$$

The latent variable, z_i, is interpreted as patient i′s underlying degree of independence on assessment item GG0130A, and is a continuous variable. The error term, ε_i, is assumed to be independent and identically distributed N(0,1). The model assumes that the assessment item, g_i, because it only can take on six levels, discretizes the underlying continuous independence. It does this using thresholds: patients whose underlying independence is lower than the lowest threshold, α_1, are coded as most dependent and given a score of 1; patients whose level of dependence is a bit higher, higher than the lowest threshold α_1 but lower than the second lowest threshold α_2, achieve a score of 2 on this item. This proceeds until we are considering patients whose independence is higher than the highest threshold, α_5, who receive a score of 6.

We compute the imputed value of g_i as

$$\widehat{g_i} = \Pr(z_i \leq \alpha_1) + 2 * \Pr(\alpha_1 < z_i \leq \alpha_2) + 3 * \Pr(\alpha_2 < z_i \leq \alpha_3) + 4 * \Pr(\alpha_3 < z_i \leq \alpha_4) + 5 * \Pr(\alpha_4 < z_i \leq \alpha_5) + 6 * \Pr(z_i > \alpha_5) \quad (8)$$

Step 5: Repeat Steps 1 - 4 for Eating (GG0130A) at discharge, replacing the word "admission" with the word "discharge" in Steps 1 - 4.

Step 6: Repeat Steps 1 - 5 for each GG item included in the observed discharge function score (Section 3.4.1), as above, replacing the Eating (GG0130A) item with each successive GG item in Steps 1-5. For Wheel 50 Feet with 2 Turns (GG0170R), use only the sample of stays that satisfies the conditions in Step 2a. For Walk 10 Feet (GG0170I) and Walk 50 Feet with 2 Turns (GG0170J), use only the sample of stays that satisfies the conditions in Step 2b.

3.6 Risk Adjustment

The purpose of risk adjustment is to account for differences across IRF patients that affect their functional status. Risk adjustment creates an individualized expectation for discharge function score for each stay that controls for admission functional status, age, and clinical characteristics. This ensures that each stay is measured against an expectation that is calibrated to the patient's individual circumstances when determining the numerator for each IRF. See the Appendix for risk adjustment model results.

3.6.1 Statistical Risk Model

The statistical risk model is an ordinary least squares linear regression model, which estimates the relationship between discharge function score and a set of risk adjustors. Observed discharge function score is determined for each IRF stay, incorporating imputed item scores when NA codes are encountered. The risk adjustment model is run on all IRF stays to determine the model intercept (β_0) and risk adjustor coefficients ($\beta_1, \ldots, \beta_n$). Expected discharge function scores are calculated by applying the regression equation to each IRF stay.

$$Expected\ Discharge\ Function\ Score = \beta_0 + \beta_1 x_1 + \cdots + \beta_n x_n \quad (9)$$

where $x_1 - x_n$ are the risk adjustors.

3.6.2 Variables

This section contains a listing of covariates groups used to calculate the risk-adjusted discharge function scores. Information on the covariates were obtained from the IRF-PAI data.

- *Age Category*

Age was calculated as the difference between the admission date of the IRF stay and the beneficiary's date of birth.

- *Admission Function Score*

  Admission function score is the sum of admission scores for function items included in the discharge score (Section 3.4.1) and can range from 10-60, with a higher score indicating greater independence. NAs in the admission item scores are treated the same way as NAs in the discharge item scores, with NAs replaced with imputed scores (Steps 1-2 in Section 3.4.1). Walking items and wheeling item are used in the same manner as in the discharge score (Step 3 in Section 3.4.1). Admission score squared is also included as a risk adjustor.

- *Primary Diagnosis Group*

  Primary diagnosis is the principal reason for admitting the patient into IRF care.

- *Interaction between Primary Diagnosis Group and Admission Function Score*

  These covariates are the admission function score multiplied by each primary diagnosis indicator.

- *Prior surgery*

  This covariate captures whether or not the patient had prior surgery.

- *Prior Function/Device Use*

  These covariates capture patient's functional status prior to the stay.

- *Pressure Ulcers*

  These covariates capture the presence of pressure ulcer(s) at different stages.

- *Cognitive Function*

  These covariates capture the patient's cognitive function by assessing whether or not the patient's mental status at admission is impaired, and if impaired, at what level.

- *Communication impairment*

  These covariates capture the patient's communication function, and indicate whether or not the patient's communication status at admission is impaired, and if impaired, at what level.

- *Incontinence*

  These covariates indicate the patient's level of bladder and bowel incontinence.

- *Nutritional Status*

  These covariates indicate patient's swallowing ability at IRF admission and patient's body mass index.

- *History of Falls*

  This covariate indicates a history of falls prior to the IRF admission.

- *HCC Comorbidities*

  Comorbidities are obtained from Item 24 in the IRF-PAI. Comorbidities are grouped using CMS Hierarchical Condition Categories (HCC) software version 24.

3.7 Measure Calculation

The Discharge Function Score is the proportion of IRF stays where patients achieve an expected discharge function score at discharge. A higher score indicates better performance in functional outcomes. For each IRF stay, the observed discharge function score (Section 3.4.1) and the expected discharge function score (Section 3.4.1) are determined. For each IRF, the Discharge Function Score is the proportion of stays where the observed discharge function score is larger than or equal to the risk-adjusted expected function score.

3.7.1 Steps Used in Calculation

Step1: Calculate the observed discharge function score as described in Section 3.4.1, incorporating imputed item scores (Section 3.5).

Step 2: Identify excluded IRF stays using the criteria mentioned in Section 3.3.2.

Step 3: Calculate the expected discharge function score. For each IRF stay, use the intercept and regression coefficients to calculate the expected discharge function score using the formula mentioned in Section 3.6. Note that any expected discharge function score greater than the maximum (i.e., 60) would be recoded to the maximum score.

Step 4: Calculate the difference in observed and expected discharge function scores. For each IRF stay which does not meet the exclusion criteria, compare each patient's observed discharge function score (Step 1) and expected discharge function score (Step 3) and classify the difference as one of the following:

- Observed discharge score is equal to or higher than the expected discharge score.
- Observed discharge score is lower than the expected discharge score.

Step 5: Determine the denominator count. Determine the total number of IRF stays with an IRF-PAI target date in the measure target period, which do not meet the exclusion criteria.

Step 6: Determine the numerator count. The numerator for this quality measure is the number of IRF stays in which the observed discharge score is the same as or higher than the expected discharge score, as determined in Step 4.

Step 7: Calculate the facility-level discharge function percent. Divide the facility's numerator count (Step 6) by its denominator count (Step 5) to obtain the facility-level discharge function percent, then multiply by 100 to obtain a percent value.

Step 8: Round the percent value to two decimal places. If the digit in the third decimal place is 5 or greater, add 1 to the second decimal place; otherwise, leave the second decimal place unchanged. Drop all digits following the second decimal place.

4 MEASURE TESTING

4.1 Reportability

Reportability testing examines the total number and proportion of stays that would have at least 20 eligible stays for the Discharge Function Score measure in the reporting period. In FY2021, 1,107 out of a total of 1,124 IRFs (98.5%) met this threshold. This indicates high reportability and usability of the measure.

Table 3. Publicly Reportable IRFs, FY2021

Number of IRFs with ≥ 20 stays	Percentage of IRFs with ≥ 20 stays
1,107	98.5%

4.2 Variability

Variability testing summarizes the distribution of the facility-level final Discharge Function Scores. In FY2021, the mean final score among IRFs with at least 20 stays was 54.2% (median: 55.5%, IQR: 42.6% - 66.4%). Final scores ranged from a minimum of 2.3% to a maximum of 95.9%. This wide variation indicates there is a performance gap in Discharge Function Scores across IRFs.

Table 4. Facility-Level Distribution of Discharge Function Scores

N	Mean Score	Std dev.	Minimum	25th percentile	50th percentile	75th percentile	Maximum
1,107	54.2%	16.8%	2.3%	42.6%	55.5%	66.4%	95.9%

4.3 Reliability

The split-half reliability test examined agreement between two Discharge Function Scores for a facility based on randomly-split, independent subsets of stays in the same measurement period. Good agreement between the two measure scores calculated in this manner provides evidence that the measure is capturing an attribute of the facility (quality of care) rather than the patient stays (case-mix). For IRFs with at least 20 eligible stays in FY2021, each provider's stays were randomly divided into halves, thus ensuring that patient stays were evenly distributed across the split-halves. Provider measure scores for each split-half sample were calculated. The Shrout-Fleiss intraclass correlation coefficient (ICC (2, 1)) was calculated between the split-half scores to measure reliability, applying the Spearman-Brown correction.[12]

[12] McGraw, K. O., & Wong, S. P. Forming inferences about some intraclass correlation coefficients. Psychological methods, 1996, 1(1), 30.

The intraclass correlation coefficient for IRFs with more than 20 eligible stays was 0.95, which indicates excellent reliability.[13]

4.4 Validity

This section reviews validity tests conducted to support the Discharge Function measure. Section 4.4.1 reports results that support the validity of measure scores. Section 4.4.2 describes analyses validating the imputation model results.

4.4.1 Measure Scores

To evaluate the validity of measure scores, convergent validity with other IRF QRP measures, face validity, and risk adjustment model performance were assessed. The following subsections describe comparisons with other measures; webinars convened to gather expert, patient, and caregiver perspectives; and risk adjustment model calibration and fit analyses.

Convergent Validity

To evaluate convergent validity, the relationships between the Discharge Function Score measure and related IRF QRP measures were examined. Using Spearman's rank correlation, the Discharge Function Score measure was compared to claims-based measure Discharge to Community (DTC) and to assessment-based functional outcome measures (Change in Self-Care Score, Discharge Self-Care Score, Change in Mobility Score, and Discharge Mobility Score). The analysis used FY2021 data from providers with at least 20 stays. As shown in Table 5, the Discharge Function measure was positively correlated with DTC (0.24) and each of the functional outcome measures: Change in Self-Care Score (0.85), Discharge Self-Care Score (0.89), Change in Mobility Score (0.89), and Discharge Mobility Score (0.91). All results were statistically significant ($p<0.01$). These results matched expectations. Higher functional status corresponds with higher likelihood of community discharge.[14] Since the other functional outcome measures use overlapping but not identical GG items and a different method for handling NA codes, the Discharge Function scores should correlate well but not perfectly with the in-use functional outcome measures.

[13] Koo T.K. & Li M.Y. A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. Journal of Chiropractic Medicine, 2016, 15(2), 155-163.
[14] Minor M, Jaywant A, Toglia J, Campo M, O'Dell MW. Discharge Rehabilitation Measures Predict Activity Limitations in Patients with Stroke Six Months after Inpatient Rehabilitation. Am J Phys Med Rehabil. 2021 Oct 20. doi: 10.1097/PHM.0000000000001908. Epub ahead of print. PMID: 34686630.

Table 5. Correlations between Discharge Function Score and Other Publicly Reported Measures

Measure	Spearman's Correlation	P value
Discharge to Community–PAC IRF QRP (NQF #3479)	0.24	<0.01
IRF Functional Outcome Measure: Change in Self-Care Score for Medical Rehabilitation Patients (NQF #2633)	0.85	<0.01
IRF Functional Outcome Measure: Discharge Self-Care Score for Medical Rehabilitation Patients (NQF #2635)	0.89	<0.01
IRF Functional Outcome Measure: Change in Mobility Score for Medical Rehabilitation Patients (NQF #2634)	0.89	<0.01
IRF Functional Outcome Measure: Discharge Mobility Score for Medical Rehabilitation Patients (NQF #2636)	0.91	<0.01

Face Validity

To assess face validity of the Discharge Function Score measure, two Technical Expert Panel (TEP) meetings (July 2021 and January 2022), as well as a Patient and Family Engagement Listening Session, were convened. TEP members showed strong support for the face validity of this measure. Though a vote was not taken at the meeting, the TEP agreed with the conceptual and operational definition of the measure. Panelists reviewed the validity analyses described herein and agreed they demonstrated measure validity.

The Patient and Family Engagement Listening Session demonstrated that the measure concept resonates with patients and caregivers. Participants' views of self-care and mobility were aligned with the functional domains captured by the measure, and they found them to be critical aspects of care. Participants emphasized the importance of measuring functional outcomes and were specifically interested in metrics that show how many patients discharged from particular facilities made improvements in self-care and mobility.

Risk Adjustment Model Performance

The risk adjustment model is an ordinary least squares (OLS) linear regression. We assessed risk adjustment model calibration and fit using FY2021 data. A well-calibrated model demonstrates good predictive ability to distinguish high-risk from low-risk patients. To assess risk adjustment model calibration, the ratios of observed-to-predicted discharge function score across eligible stays by decile of predicted discharge function score (risk) were calculated. The average ratios of observed-to-predicted scores for each risk decile ranged from 0.99 to 1.00, which suggested good calibration across the range of patients without evidence of concerning under- or over-estimation. Model fit was analyzed using adjusted R-squared to determine if the risk adjustment model can accurately predict discharge function while controlling for patient case-mix. The adjusted R-squared value was 0.51, which suggests good model discrimination.

4.4.2 Imputation Model

This section discusses the validity testing results of the imputation models used to estimate missing item scores. Validity testing included (1) assessments of model results and (2) calculations of bias and error of imputed item scores.

Model Results

To assess the validity of the imputation models, model fit and face validity of model coefficients were evaluated. The C-statistic is a measure of model discrimination that determines the probability that predicting the outcome is better than chance. The C-statistic can range from 0.5 to 1. Using FY2021 data, the C-statistic averaged 0.93 and ranged from 0.77 to 0.99 across the imputation models for each item at both admission and discharge (see Table A-2). These results suggests good model discrimination across all imputation models.

The face validity of model results was assessed by reviewing model coefficients. For each item at both admission and discharge, imputation models produced sensible coefficients. Worse health conditions generally predicted lower item scores, as did prior functional status. Coefficients on related GG items were positively predictive, and larger for GG items more closely related to the item being imputed (e.g., bed mobility items were generally more predictive for a bed mobility item imputation model than transfer or ambulation items).[15]

Bias and Mean Squared Error

A bootstrapping method was used to measure bias and mean squared error (MSE) in the imputation method. Bias measures the average amount by which the imputed value differs from the true value. Bias is signed, with a positive amount meaning that the imputed values were higher, on average, than the true values. MSE measures how far away the method is, on average from the truth. It is unsigned and can be positive even if bias is zero. The absolute size of bias is an inverse measure of accuracy, while the size of MSE is an inverse measure of the combination of precision and accuracy. The goal of the bootstrapping method was to determine how similar imputed values were to the true item score. This similarity could not be measured directly since the true value of the measure score was unknown in the case of the individuals for whom imputation was necessary (imputation was needed precisely because the missing values prevented calculating the measure score for these individuals). Therefore, a bootstrapping strategy was implemented using the following steps to assess the accuracy of the statistical imputation method:

Step 1: Identified observations from the original sample with no NAs recorded across all items needed for measure calculation.

[15] Detailed model results are available upon request.

Step 2: Generated a bootstrap sample that draws from the no-NA observations until there were as many observations in the bootstrap sample as the original sample. A stratified random sampling algorithm was used. The first stratum of each bootstrap sample consisted of no-NA observations. This stratum had the same number of observations as there were no-NA observations in the original data. This stratum of the bootstrap sample was filled by simple random sampling from the no-NA observations.

To fill the bootstrap sample observations corresponding to the observations from the original data having NAs, it was not possible to use simple random sampling. This is because the distribution of clinical and function characteristics was different between observations with and without NAs. Therefore, the sampling to fill the bootstrap sample for these observations was done using a stratification method which matched observations with NA to similar observations without NA.

Therefore, ten additional strata were filled corresponding to the observations from the original data with NAs. These strata were defined by the deciles of a predicted score estimated, as described in Section 3.5. Bootstrap observations corresponding to the observations with NAs were chosen by simple random sampling within each of these strata.

Step 3: Created two copies of this sample.

a) One copy served as the gold standard source of truth because all observations in the bootstrap sample were sampled from no-NA observations.

b) In the other copy, NAs were imposed on some of the GG items. This was done in a way which preserved both the pattern of NAs within the data and the pattern of clinical characteristics among NA observations. NAs were imposed by randomly selecting observations from the original data which i) had NAs and ii) were in the same stratum (see Step 2) as the corresponding target observation in the second copy. The GG items which were missing in the sampled observation were made missing in the target observation.

Step 4: In the second copy produced in Step 3b, imputed values for the NAs imposed onto the bootstrap sample were generated. For comparison, applied "recode to 1" method and calculated resulting measure scores.

Step 5: Calculated bias and mean-squared error of the imputation method by comparing observation by observation to the measure scores produced from the gold standard copy (Step 3a).

Step 6: Repeated Steps 2-5 many times. Reported average bias/mean-squared error across iterations/bootstrap replications.

Bias and MSE were compared between statistical imputation and the current method for in-use measures, which recodes all NAs to 1. Using this bootstrapping method, statistical imputation resulted in lower levels of bias (-0.39 at admission; -0.07 at discharge) and MSE (2.17 at admission; 0.50 at discharge) compared to the bias (-1.43 at admission; -0.51 at discharge) and MSE (6.99 at admission; 2.58 at discharge) produced from the recode approach, which supports the validity of the statistical imputation method.

APPENDIX

Table A-1. Discharge Function Score Measure Risk Adjustment: Linear Regression Model Results, FY2021

Covariate	Number of Stays	Percent of Stays	Average Observed Score	Estimate	P-value
Age Group < 35 years	1,368	0%	49.70	-0.19	0.35
Age Group 35 - 44 years	4,071	1%	50.44	0.63	0.00
Age Group 45 - 54 years	11,273	3%	50.76	0.71	0.00
Age Group 55 - 64 years	34,854	8%	50.43	0.26	0.00
Age Group 75 - 84 years	151,325	34%	49.10	-0.72	0.00
Age Group 85 - 90 years	56,253	13%	47.46	-1.56	0.00
Age Group > 90 years	23,660	5%	45.89	-2.43	0.00
Admission Score - continuous form	-	-	.	1.28	0.00
Admission Score - squared form	-	-	.	-0.02	0.00
Primary Diagnosis Group: Stroke	98,225	22%	46.11	-13.83	0.00
Primary Diagnosis Group: Non-Traumatic Brain Dysfunction	29,935	7%	48.00	-8.56	0.00
Primary Diagnosis Group: Traumatic Brain Dysfunction	18,265	4%	48.62	-6.06	0.00
Primary Diagnosis Group: Non-Traumatic Spinal Cord Dysfunction	19,253	4%	48.88	-10.53	0.00
Primary Diagnosis Group: Traumatic Spinal Cord Dysfunction	4,236	1%	43.90	-14.22	0.00
Primary Diagnosis Group: Progressive Neurological Conditions	9,175	2%	45.71	-9.18	0.00
Primary Diagnosis Group: Other Neurological Conditions	51,174	12%	50.78	-6.43	0.00
Primary Diagnosis Group: Fractures and Other Multiple Trauma	59,700	14%	49.83	-4.56	0.00
Primary Diagnosis Group: Amputation	12,707	3%	49.38	-9.29	0.00
Primary Diagnosis Group: Other Orthopedic Conditions	32,505	7%	51.08	-6.59	0.00
Primary Diagnosis Group: Debility, Cardiorespiratory Conditions	86,721	20%	51.79	-6.91	0.00
Primary Diagnosis Group: Medically Complex Conditions	6,415	1%	51.16	-7.42	0.00
Interaction of Admission Score and Primary Diagnosis Group: Stroke	-	-	.	0.34	0.00
Interaction of Admission Score and Primary Diagnosis Group: Non-Traumatic Brain Dysfunction	-	-	.	0.20	0.00
Interaction of Admission Score and Primary Diagnosis Group: Traumatic Brain Dysfunction	-	-	.	0.12	0.00

Covariate	Number of Stays	Percent of Stays	Average Observed Score	Estimate	P-value
Interaction of Admission Score and Primary Diagnosis Group: Non-Traumatic Spinal Cord Dysfunction	-	-	.	0.24	0.00
Interaction of Admission Score and Primary Diagnosis Group: Traumatic Spinal Cord Dysfunction	-	-	.	0.36	0.00
Interaction of Admission Score and Primary Diagnosis Group: Progressive Neurological Conditions	-	-	.	0.18	0.00
Interaction of Admission Score and Primary Diagnosis Group: Other Neurological Conditions	-	-	.	0.18	0.00
Interaction of Admission Score and Primary Diagnosis Group: Fractures and Other Multiple Trauma	-	-	.	0.10	0.00
Interaction of Admission Score and Primary Diagnosis Group: Amputation	-	-	.	0.19	0.00
Interaction of Admission Score and Primary Diagnosis Group: Other Orthopedic Conditions	-	-	.	0.15	0.00
Interaction of Admission Score and Primary Diagnosis Group: Debility, Cardiorespiratory Conditions	-	-	.	0.18	0.00
Interaction of Admission Score and Primary Diagnosis Group: Medically Complex Conditions	-	-	.	0.17	0.00
Prior Surgery - Surgical	195,243	44%	50.25	0.08	0.00
Prior Functioning: Indoor Ambulation - Dependent	9,227	2%	38.79	-2.70	0.00
Prior Functioning: Indoor Ambulation -Some Help	53,574	12%	43.65	-1.17	0.00
Prior Functioning: Stair Negotiation - Dependent	57,227	13%	45.41	-0.47	0.00
Prior Functioning: Stair Negotiation - Some Help	64,730	15%	46.35	-0.17	0.00
Prior Functioning: Cognition - Dependent	6,285	1%	38.29	-1.15	0.00
Prior Mobility Device/Aid - Walker	191,174	43%	48.20	-0.11	0.00
Prior Mobility Device/Aid - Manual Wheelchair or Motorized Wheelchair and/or Scooter	62,930	14%	45.13	-0.86	0.00
Prior Mobility Device/Aid - Mechanical Lift	3,406	1%	39.86	-2.17	0.00
Prior Mobility Device/Aid - Orthotics/Prosthetics	4,752	1%	48.31	-0.25	0.02
Prior Functioning: Self Care Some Help	86,213	20%	44.22	-1.75	0.00
Prior Functioning: Self Care Dependent	3,270	1%	32.83	-3.99	0.00

Covariate	Number of Stays	Percent of Stays	Average Observed Score	Estimate	P-value
Stage 2 Pressure Ulcer – Admission	20,494	5%	45.09	-1.00	0.00
Stage 3, 4 or Unstageable Pressure Ulcer/Injury - Admission	29,427	7%	43.85	-1.58	0.00
Cognitive Function: Brief Interview for Mental Status score - Admission - Moderately Impaired	97,793	22%	47.47	-0.95	0.00
Cognitive Function: Brief Interview for Mental Status score - Admission - Severely Impaired	37,829	9%	41.61	-1.91	0.00
Communication Impairment - Admission - Moderate to Severe	50,251	11%	40.45	-1.63	0.00
Communication Impairment - Admission - Mild	158,764	36%	46.93	-0.64	0.00
Bladder Incontinence - Admission - Indwelling urinary catheter	37,753	9%	44.61	-1.71	0.00
Bladder Incontinence - Admission - Less than Daily, Daily, Always Incontinent	148,423	34%	44.32	-1.54	0.00
Bowel Incontinence - Admission - Always Incontinent	40,706	9%	38.30	-3.01	0.00
Bowel Incontinence - Admission - Less than Daily, Daily	63,427	14%	45.03	-1.31	0.00
Health Conditions - Admission - History of Falls	219,067	50%	48.77	-0.27	0.00
Swallowing Ability - Admission - Tube/Parenteral Feeding	12,766	3%	40.36	-0.96	0.00
Swallowing Ability at Admission -- modified food consistency	92,327	21%	44.64	-0.68	0.00
Total Parenteral Nutrition	1,322	0%	46.01	-1.32	0.00
High BMI (BMI > 50)	7,545	2%	48.80	-0.48	0.00
Low BMI	23,014	5%	47.90	-0.57	0.00
Major Infections: Septicemia - Sepsis - Systemic Inflammatory Response Syndrome / Shock (HCC2)	8,430	2%	48.39	0.05	0.57
Opportunistic Infections (HCC6)	1,367	0%	48.98	-0.74	0.00
Metastatic Cancer and Acute Leukemia (HCC8)	6,947	2%	48.52	-1.66	0.00
Lung and Other Severe Cancers (HCC9)	7,317	2%	49.52	-0.76	0.00
Lymphoma and Other Cancers (HCC10)	6,152	1%	49.09	-0.65	0.00
Diabetes with Chronic Complications (HCC18)	122,910	28%	48.93	-0.18	0.00
Diabetes without Complication (HCC19)	82,021	19%	49.20	-0.12	0.00
Other Significant Endocrine and Metabolic Disorders (HCC23)	17,582	4%	48.76	-0.35	0.00
Intestinal Obstruction/Perforation (HCC33)	5,201	1%	48.44	-0.45	0.00

Covariate	Number of Stays	Percent of Stays	Average Observed Score	Estimate	P-value
Bone/Joint/Muscle Infections/Necrosis (HCC39)	5,582	1%	48.16	-0.93	0.00
Severe Hematological Disorders (HCC46)	2,027	0%	49.11	-0.50	0.00
Dementia With Complications (HCC51)	4,219	1%	42.32	-1.77	0.00
Dementia Without Complications (HCC52)	34,410	8%	43.81	-1.50	0.00
Mental Health Disorders: Schizophrenia (HCC57)	3,330	1%	49.15	-0.10	0.46
Major Depressive, Bipolar, and Paranoid Disorders (HCC58)	549	0%	44.51	-1.03	0.00
Reactive and Unspecified Psychosis (HCC59)	25,818	6%	49.23	-0.04	0.43
Personality Disorders (HCC60)	519	0%	50.27	-0.63	0.06
Tetraplegia (HCC70)	1,616	0%	41.10	-2.70	0.00
Paraplegia (HCC71)	1,722	0%	45.01	-1.72	0.00
Spinal Cord Disorders/Injuries (HCC72)	3,272	1%	47.54	-0.76	0.00
Amyotrophic Lateral Sclerosis and Other Motor Neuron Disease (HCC73)	169	0%	43.65	-2.51	0.00
Cerebral Palsy (HCC74)	857	0%	45.45	-1.78	0.00
Muscular Dystrophy (HCC76)	243	0%	46.50	-2.16	0.00
Multiple Sclerosis (HCC77)	2,857	1%	47.64	-1.13	0.00
Parkinson's and Huntington's Diseases (HCC78)	12,380	3%	45.02	-1.50	0.00
Seizure Disorders and Convulsions (HCC79)	25,447	6%	47.04	-0.51	0.00
Angina Pectoris (HCC88)	1,929	0%	50.18	0.15	0.36
Atherosclerosis of the Extremities with Ulceration or Gangrene (HCC106)	1,755	0%	47.54	-1.00	0.00
Aspiration and Specified Bacterial Pneumonias (HCC114)	6,086	1%	45.20	-0.29	0.00
Pneumococcal Pneumonia, Empyema, Lung Abscess (HCC115)	1,330	0%	49.46	0.36	0.08
Proliferative Diabetic Retinopathy and Vitreous Hemorrhage (HCC122)	54	0%	46.68	-1.76	0.08
Dialysis and Chronic Kidney Disease - Stage 5: Dialysis Status (HCC134), Chronic Kidney Disease, Stage 5 (HCC136)	23,611	5%	48.39	-1.44	0.00
Acute Renal Failure (HCC135)	50,027	11%	48.27	-0.52	0.00
Chronic Kidney Disease, Severe (Stage 4) (HCC137)	10,190	2%	48.68	-0.18	0.02
Chronic Kidney Disease, Moderate (Stage 3) (HCC138)	52,696	12%	49.08	-0.02	0.59

Covariate	Number of Stays	Percent of Stays	Average Observed Score	Estimate	P-value
Chronic Ulcer of Skin, Excluding Pressure Ulcer (HCC161)	9,900	2%	48.11	-0.36	0.00
Severe Skin Burn (HCC162)	73	0%	51.09	1.25	0.15
Major Head Injury (HCC167)	2,185	0%	48.03	-0.65	0.00
Hip Fracture/Dislocation (HCC170)	1,026	0%	47.23	-0.95	0.00
Amputations: Traumatic Amputations and Complications (HCC173)	229	0%	48.85	-0.78	0.11
Complication of Specified Implanted Device or Graft (HCC176)	2,669	1%	46.82	-0.67	0.00
Major Organ Transplant or Replacement Status (HCC186)	3,780	1%	51.17	-0.63	0.00
Amputation Status, Lower Limb/ Amputation Complications (HCC189)	8,248	2%	49.16	-0.15	0.09
Cerebral Hemorrhage (HCC99); Ischemic or Unspecified Stroke (HCC100); Hemiplegia/Hemiparesis (HCC103);	81,859	19%	45.26	-1.11	0.00
Intercept	.	.	.	32.62	0.00

Table A-2. C-Statistics for Imputation Models across GG Items at Admission and Discharge, FY2021

Item	Description	Assessment Timing	C-Statistic
GG0130A	Eating	Admission	0.80
		Discharge	0.91
GG0130B	Oral Hygiene	Admission	0.77
		Discharge	0.91
GG0130C	Toileting Hygiene	Admission	0.87
		Discharge	0.92
GG0170A	Roll left/right	Admission	0.90
		Discharge	0.96
GG0170C	Lying to sit - bed	Admission	0.95
		Discharge	0.98
GG0170D	Sit to stand	Admission	0.95
		Discharge	0.97
GG0170E	Chair to bed trans.	Admission	0.96
		Discharge	0.97
GG0170F	Toilet trans.	Admission	0.91
		Discharge	0.94
GG0170I	Walk 10'	Admission	0.90
		Discharge	0.98
GG0170J	Walk 50'	Admission	0.97
		Discharge	0.99
GG0170R	Wheel 50'	Admission	0.90
		Discharge	0.96