Discharge Function Score for Long-Term Care Hospitals (LTCHs)

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 long-term care hospital (LTCH) 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 LTCH Quality Reporting Program (QRP) under the *Quality Measure & Assessment Instrument Development & Maintenance & QRP Support* contract (75FCMC18D0015/Task Order 75FCMC19F0003).

Measuring functional status of LTCH patients can provide valuable information about a LTCH's quality of care. Physical function predicts several outcomes including successful discharge to the community and re-hospitalization rates.[3,4] Several studies have reported that LTCH care can improve patients' motor function at discharge for patients with various clinical circumstances, including recent ventilation and cognitive impairment.[5,6] Providers can intervene to improve patients' functional outcomes by adopting a patient-centered care plan that accounts for each patient's unique circumstances.[3,7]

The Discharge Function Score measure determines how successful each LTCH is at achieving an expected level of functional ability for its patients at discharge. An expectation for discharge function score is built for each LTCH stay by accounting for patient characteristics that impact their functional status. The final Discharge Function Score for a given LTCH is the proportion of that LTCH's stays where a patient's observed discharge score meets or exceeds their expected discharge score. LTCHs with low scores are not producing the functional gains

[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] Dubin R, Veith JM, Grippi MA, McPeake J, Harhay MO, Mikkelsen ME. Functional Outcomes, Goals, and Goal Attainment among Chronically Critically Ill Long-Term Acute Care Hospital Patients. Ann Am Thorac Soc. 2021;18(12):2041-2048. doi:10.1513/AnnalsATS.202011-1412OC
[4] 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.
doi:10.1097/CCM.0000000000005089
[5] Mayer KP, Pastva AM, Du G, et al. Mobility Levels With Physical Rehabilitation Delivered During and After Extracorporeal Membrane Oxygenation: A Marker of Illness Severity or an Indication of Recovery?. Phys Ther. 2022;102(3):pzab301. doi:10.1093/ptj/pzab301
[6] Lane NE, Stukel TA, Boyd CM, Wodchis WP. Long-Term Care Residents' Geriatric Syndromes at Admission and Disablement Over Time: An Observational Cohort Study. J Gerontol A Biol Sci Med Sci. 2019;74(6):917-923. doi:10.1093/gerona/gly151
[7] 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.

that they could be for a larger share of their patients. The measure provides actionable feedback to LTCHs 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 LTCH setting. The Discharge Function Score measure adds value to the LTCH QRP function measure portfolio by using specifications that allow for better comparisons across post-acute are (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 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

LTCH

2.4 Data Source

Long-Term Care Hospital (LTCH) Continuity Assessment Record and Evaluation (CARE) Data Set (LCDS)

2.5 Brief Description of Measure

The Discharge Function Score measure assesses the percent of LTCH patients who achieve a risk-adjusted expected function score at discharge. Functional status is measured through Section GG of LCDS 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 LTCH 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 Study Window

This measure is calculated using 12 months (four quarters) of data. All LTCH 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 LCDS. The LCDS data are collected on all patients who receive services from a long-term care hospitals. This measure is calculated entirely using administrative data. There will be no additional data collection or submission burden for LTCH providers as the data used in the measure are already collected on the LCDS.

3.3 Denominator

The denominator is the total number of LTCH stays with a LCDS record in the measure target period, which do not meet the exclusion criteria.

3.3.1 Stay Construction

LTCH stays require a matched pair of Admission and Discharge assessments that are matched based on a unique combination of Provider Internal ID, Patient Internal ID, and Admission Date. In the case of multiple discharge assessments (where there is more than one discharge record for the same patient-provider-date combination), the discharge record with the highest discharge Reason For Assessment (RFA) value is used, i.e., in the following order: expired record, unplanned discharge, planned discharge. Any unmatched records, stays with invalid dates, or overlapping stays are excluded.

3.3.2 Eligible Stays

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

- Patients with incomplete stays. Patients with incomplete stays include patients who have unplanned discharge or are discharged against medical advice; patients who died during the LTCH stay; patients with discharge to hospital emergency department, short-stay acute hospital, psychiatric hospital/unit, or long-term care hospital; and patients with a length of stay of 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 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 LTCH patients may have different patterns of care than adult patients.

- Patient is discharged to hospice.

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

- Patient is in a coma, persistent vegetative state, or locked-in syndrome, or has complete tetraplegia.

  Rationale: These patients are excluded because they may have limited or less predictable mobility improvement with the selected items.

3.4 Numerator

The numerator is the number of patients in a LTCH 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 LTCH 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 the in-use LTCH Functional Outcome Measure: Change in Mobility Among Long-Term Care Hospital Patients Requiring Ventilator Support (NQF #2632). The method used in previously specified 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 LTCH 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_{10}) 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^{**}).

$$Function\ items: G \in \{g_2, \dots, g_{10}\} \tag{1}$$

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

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

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

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

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

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.

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

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

The model we estimate for g_1, GG0130A, is

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

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

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

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

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 LTCH 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 LTCH. 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 LTCH stay, incorporating imputed item scores when NA codes are encountered. The risk adjustment model is run on all LTCH 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 LTCH stay.

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

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 LCDS data.

- *Age Category*

  Age was calculated as the difference between the admission date of the LTCH 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 Medical Condition Category*

Primary medical condition is the principal reason for admitting the patient into LTCH care.

- *Interaction between Primary Medical Condition Category and Admission Function Score*

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

- *Ventilator Status*

  This covariate indicates whether the patient required ventilator support on admission.

- *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.

- *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 total parenteral nutrition status at LTCH admission and patient's body mass index.

- *Comorbidities*

  Comorbidities are obtained from Section I in the LCDS.

3.7 Measure Calculation

The Discharge Function Score is the proportion of LTCH stays with patients achieving an expected discharge function score at discharge. A higher score indicates better performance in functional outcomes. For each LTCH stay, the observed discharge function score (Section 3.4.1) and the expected discharge function score (Section 3.4.2) are determined. For each LTCH, 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 LTCH stays using the criteria mentioned in Section 3.3.2.

Step 3: Calculate the expected discharge function score. For each LTCH stay: use the intercept and regression coefficients to calculate the expected discharge function score using the formula described in Section 3.4.2. 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 LTCH 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 LTCH stays with discharge dates 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 LTCH 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, 334 out of a total of 341 LTCHs (97.9%) met this threshold. This indicates high reportability and usability of the measure.

Table 3. Publicly Reportable LTCHs, FY2021

Number of LTCHs with ≥ 20 stays	Percentage of LTCHs with ≥ 20 stays
334	97.9%

4.2 Variability

Variability testing summarizes the distribution of the facility-level final Discharge Function Scores. In FY2021, the mean final score among LTCHs with at least 20 stays was 48.7% (median: 48.1%, IQR: 37.7% - 59.6%). Final scores ranged from a minimum of 5.2% to a maximum of 90.8%. This wide variation indicates there is a performance gap in Discharge Function Scores across LTCHs.

Table 4. Facility- Level Distribution of Discharge Function Scores

N	Mean Score	Std dev.	Minimum	25th percentile	50th percentile	75th percentile	Maximum
334	48.7%	15.4%	5.2%	37.7%	48.1%	59.6%	90.8%

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 LTCHs 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.[8]

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

The intraclass correlation coefficient for LTCHs with more than 20 eligible stays was 0.96, which indicates excellent reliability.[9]

4.4 Validity

This section reviews validity tests conducted to support the Discharge Function Score 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 LTCH 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 LTCH QRP measures were examined. Using Spearman's rank correlation, the Discharge Function Score measure was compared to claims-based measures of Discharge to Community (DTC), Potentially Preventable Readmissions within 30-Days Post-Discharge (PPR), and Medicare Spending Per Beneficiary (MSPB) 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. Higher functional status corresponds with higher likelihood of community discharge and lower rates of re-hospitalizations.[10] As expected, this measure demonstrated positive correlation with DTC (0.45) and negative correlation with PPR (-0.16). Because higher functioning patients are likely to have lower levels of medical complexity, it follows that their stays would cost less. As expected, this measure had a negative correlation with MSPB (-0.20). Additionally, as expected, since the LTCH QRP mobility functional outcome measure uses overlapping but not identical GG items, a different method for handling missing data, and is subset to the ventilated population, scores for the Discharge Function measure correlated well but not perfectly with the Functional Outcome measure: Change in Mobility for Ventilated LTCH Patients (0.81). All correlation coefficients were significant ($p<0.01$).

[9] 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.

[10] Dubin R, Veith JM, Grippi MA, McPeake J, Harhay MO, Mikkelsen ME. Functional Outcomes, Goals, and Goal Attainment among Chronically Critically Ill Long-Term Acute Care Hospital Patients. *Ann Am Thorac Soc*. 2021;18(12):2041-2048. doi:10.1513/AnnalsATS.202011-1412OC

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

Measure	Spearman's Correlation	P value
Discharge to Community - PAC LTCH QRP	0.45	<0.01
Medicare Spending Per Beneficiary – Post-Acute Care (PAC) LTCH QRP	-0.20	<0.01
Potentially Preventable 30-Day Post-Discharge Readmission Measure – LTCH QRP	-0.16	<0.01
LTCH Functional Outcome Measure: Change in Mobility Among Long-Term Care Hospital Patients Requiring Ventilator Support (NQF #2632)	0.81	<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 0.98 to 1.06, 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 0.65, 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 0.95 and ranged from 0.84 to 0.98 across the imputation models for each item at both admission and discharge (see Table A-2). These results suggest 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).[11]

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.

[11] 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, which recodes all NAs to 1. Using this bootstrapping method, statistical imputation resulted in lower levels of bias (-0.02 at admission; -0.24 at discharge) and MSE (5.98 at admission; 3.12 at discharge) compared to the bias (-2.73 at admission; -1.54 at discharge) and MSE (20.21 at admission; 9.80 at discharge) produced from the current recode, 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 18 - 54	22,064	23%	39.95	3.19	0.00
Age Group 55 - 64	22,614	24%	35.89	1.75	0.00
Age Group 65 - 74	28,564	30%	32.72	0.00	.
Age Group 75 - 84	16,448	17%	30.35	-1.56	0.00
Age Group > 85	5,102	5%	28.24	-2.61	0.00
Admission Score	-	-	.	1.50	0.00
Admission Score Squared	-	-	.	-0.01	0.00
Primary Medical Condition Category Chronic respiratory condition (PMCC 2)	1,434	2%	27.31	-2.20	0.00
Primary Medical Condition Category Acute onset and chronic respiratory conditions (PMCC 3)	11,004	12%	29.77	-1.10	0.00
Primary Medical Condition Category Chronic cardiac condition (PMCC 4)	783	1%	35.58	-2.21	0.02
Primary Medical Condition Category Other medical condition (PMCC 5)	52,649	56%	37.06	-0.53	0.00
Interaction between Admission Score & PMCC 2	-	-	.	0.02	0.34
Interaction between Admission Score & PMCC 3	-	-	.	0.02	0.06
Interaction between Admission Score & PMCC 4	-	-	.	0.06	0.06
Interaction between Admission Score & PMCC 5	-	-	.	0.02	0.00
Bladder Incontinence - Admission - Indwelling urinary catheter	28,178	30%	28.17	-2.30	0.00
Bladder Incontinence Less than Daily, Daily, Always Incontinent	33,117	35%	27.89	-2.41	0.00
Bowel Incontinence Always Incontinent	38,025	40%	24.75	-1.30	0.00
Bowel Incontinence Less than Daily, Daily Incontinent	13,375	14%	34.76	0.33	0.00
Communication Impairment Moderate to Severe	26,192	28%	21.98	-2.11	0.00
Communication Impairment Mild	23,407	25%	30.04	-0.01	0.86
Prior Functioning: Indoor Ambulation Dependent	15,124	16%	21.23	-3.60	0.00
Prior Functioning: Indoor Ambulation Some help	17,335	18%	32.42	-1.85	0.00
Prior Device Use: Mechanical Lift	2,367	2%	19.59	-1.99	0.00

Covariate	Number of Stays	Percent of Stays	Average Observed Score	Estimate	P-value
Prior Device Use: Manual Wheelchair or Motorized and/or Scooter	10,996	12%	29.48	-1.48	0.00
Stage 2 Pressure Ulcer Presence	9,936	10%	27.05	-1.46	0.00
Stage 3, 4, or unstageable pressure ulcer/injury Presence	34,836	37%	26.63	-2.56	0.00
High BMI	4,940	5%	33.21	-2.24	0.00
Low BMI	5,937	6%	31.72	-0.62	0.00
Special Treatment: Total Parenteral Nutrition	3,414	4%	34.78	0.32	0.07
Ventilator Flag: Require Ventilator at Admission	20,509	22%	24.28	0.83	0.00
Comorbidities: Acute Renal Failure	15,433	16%	31.24	0.08	0.41
Comorbidities: Amyotrophic Lateral Sclerosis, Other Progressive Neuromuscular Disease	1,052	1%	26.08	-2.20	0.00
Comorbidities: Central Nervous System Infections, Opportunistic Infections, Bone/Joint/Muscle Infections/Necrosis	19,689	21%	35.64	0.38	0.00
Comorbidities: Dementia	4,929	5%	22.11	-1.39	0.00
Comorbidities: Diabetes Mellitus (DM)	41,680	44%	33.36	-0.35	0.00
Comorbidities: Huntington's Disease, Parkinson's Disease	1,262	1%	23.27	-1.77	0.00
Comorbidities: Dialysis and Chronic Kidney Disease, Stage 5	11,298	12%	30.56	-1.94	0.00
Comorbidities: Kidney Transplant	551	1%	35.32	-0.33	0.45
Comorbidities: Major Lower Limb Amputation	2,778	3%	31.88	-2.20	0.00
Comorbidities: Major Organ Transplant	444	0%	38.08	0.33	0.49
Comorbidities: Metastatic Cancers	1,810	2%	34.60	-0.56	0.03
Comorbidities: Multiple Sclerosis (MS)	597	1%	24.02	-3.43	0.00
Comorbidities: Other Severe Neurological Injury, Disease, or Dysfunction	6,160	6%	27.37	0.05	0.74
Comorbidities: Paraplegia	2,668	3%	28.73	-2.73	0.00
Comorbidities: Peripheral Vascular Disease (PVD) or Peripheral Arterial Disease (PAD)	7,233	8%	33.89	0.10	0.43
Comorbidities: Incomplete Tetraplegia, Other Spinal Cord Disorder/Injury	2,750	3%	26.25	-1.76	0.00

Covariate	Number of Stays	Percent of Stays	Average Observed Score	Estimate	P-value
Comorbidities: Septicemia, Sepsis, Systemic Inflammatory Response Syndrome/Shock	15,543	16%	31.55	0.08	0.36
Comorbidities: Stroke, Hemiplegia or Hemiparesis	11,408	12%	24.61	-2.85	0.00
Comorbidities: Severe Cancers	4,807	5%	35.70	-0.05	0.73
Comorbidities: Severe Left Systolic/Ventricular Dysfunction	4,632	5%	31.80	0.43	0.01
Intercept	.	.	.	10.13	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	0.88
		Discharge	0.93
GG0130B	Oral Hygiene	Admission	0.84
		Discharge	0.94
GG0130C	Toileting Hygiene	Admission	0.93
		Discharge	0.93
GG0170A	Roll left/right	Admission	0.93
		Discharge	0.95
GG0170C	Lying to sit - bed	Admission	0.98
		Discharge	0.98
GG0170D	Sit to stand	Admission	0.97
		Discharge	0.97
GG0170E	Chair to bed trans.	Admission	0.98
		Discharge	0.98
GG0170F	Toilet trans.	Admission	0.98
		Discharge	0.98
GG0170I	Walk 10'	Admission	0.96
		Discharge	0.96
GG0170J	Walk 50'	Admission	0.98
		Discharge	0.98
GG0170R	Wheel 50'	Admission	0.97
		Discharge	0.96