NATIONAL QUALITY FORUM

MEASURE WORKSHEET

This document summarizes the evaluation of the measure as it progresses through NQF's Consensus Development Process (CDP). The information submitted by measure developers/stewards is included after the Brief Measure Information, Preliminary Analysis, and Pre-meeting Public and Member Comments sections.

To navigate the links in the worksheet: Click to go to the link. ALT + LEFT ARROW to return

Purple text represents the responses from measure developers.

Red text denotes developer information that has changed since the last measure evaluation review.

Brief Measure Information

NQF #: 2632

Corresponding Measures:

De.2. Measure Title: Long-Term Care Hospital (LTCH) Functional Outcome Measure: Change in Mobility Among Patients Requiring Ventilator Support

Co.1.1. Measure Steward: Centers for Medicare & Medicaid Services

De.3. Brief Description of Measure: This measure estimates the risk-adjusted change in mobility score between admission and discharge among LTCH patients requiring ventilator support at admission.

1b.1. Developer Rationale: Patients in LTCHs present with clinically complex conditions. In addition to having complex medical care needs for an extended period of time, LTCH patients often have functional limitations due to the nature of their conditions, as well as deconditioning due to prolonged bed rest and treatment requirements (for example, ventilator use). These patients are therefore at high risk for functional decline during the LTCH stay that is both condition-related and iatrogenic (i.e., related to medical treatment).

According to the Centers for Disease Control and Prevention (CDC), more than 300,000 patients receive mechanical ventilation in the United States each year (CDC 2014). These patients are at increased risk for infections, such as pneumonia and sepsis, as well as other serious complications including pulmonary edema, pulmonary embolism, and death (Esteban et al 2002; Klompas et. al 2011). These complications can lead to longer stays in the intensive care unit and hospital, increased health care costs and increased risk of disability (or death) (CDC 2014). The estimated mortality rate in patients aged 85 years and older with acute lung injury on mechanical ventilation is 60 percent (Rubenfeld et al. 2005).

A Medicare Payment Advisory Commission analysis of Medicare data found that 16 percent of LTCH patients used at least one ventilator-related service in 2012 (MedPAC 2014). In fiscal year 2012, MS-LTC-DRG 207, a diagnosis-related group that refers to respiratory diagnosis with ventilator support for 96 or more hours, represented the most frequently occurring diagnosis among LTCH patients, at 11.3 percent of all LTCH discharges (MedPAC 2014) and MS-LTC-DRG-4, a diagnosis-related group that refers to tracheostomy with ventilator support for 96 or more hours or primary diagnosis except face, mouth, and neck without major OR procedure, represented an additional 1.3 percent of all LTCH discharges. Together, the two diagnosis-related groups account for a total of nearly 18,000 discharges. Furthermore, the number of ventilated patients in LTCHs has increased. The number of patients discharged with a respiratory diagnosis with ventilator support for 96 or more hours increased 7.4 percent between 2008 and 2011 (MedPAC 2014).

Functional improvement is particularly relevant for patients who require ventilator support because these patients have traditionally had limited mobility due to cardiovascular and pulmonary instability, delirium, sedation, lack of rehabilitation therapy staff, and lack of physician referral (Zanni et. al 2010).

In describing the importance of functional status, the National Committee on Vital and Health Statistics Subcommittee on Health (2001) noted, "Information on functional status is becoming increasing essential for fostering healthy people and a health population. Achieving optimal health and well-being for Americans requires an understanding across the life space of the effects of people's health conditions on their ability to do basic activities and participate in life situations, in other words, their functional status."

This quality measure will inform LTCH providers about opportunities to improve care in the area of function and strengthen incentives for quality improvement related to patient function.

Citations

Centers for Disease Control and Prevention (CDC). Ventilator-Associated Event (VAE). January 2014. http://www.cdc.gov/nhsn/PDFs/pscManual/10-VAE_FINAL.pdf.

Esteban, A., A. Anzueto, et al. (2002). Characteristics and outcomes in adult patients receiving mechanical ventilation: a 28-day international study. JAMA 287(3): 345-355.

Klompas, M., Y. Khan, et al. (2011). Multicenter Evaluation of a Novel Surveillance Paradigm for Complications of Mechanical Ventilation. PLoS ONE 6(3): e18062.

MedPAC. Report to Congress: Medicare Payment Policy. Chapter 11: Long-term care hospital services. March 2014. http://www.medpac.gov/chapters/Mar14_Ch11.pdf.

Rubenfeld, G. D., E. Caldwell, et al. (2005). Incidence and outcomes of acute lung injury. N Engl J Med 353(16): 1685-1693.

Zanni, J. M., Korupolu, R., Fan, E., Pradhan, P., Janjua, K., Palmer, J. B., Needham, D. M. (2010). Rehabilitation therapy and outcomes in acute respiratory failure: an observational pilot project. Journal of Critical Care, 25(2), 254-262. doi: 10.1016/j.jcrc.2009.10.010

National Committee on Vital and Health Statistics Subcommittee on Health. (2001). Classifying and Reporting Functional Status.

S.4. Numerator Statement: The measure does not have a simple form for the numerator and denominator. This measure estimates the risk-adjusted change in mobility score between admission and discharge among LTCH patients requiring ventilator support at admission. The change in mobility score is calculated as the difference between the discharge mobility score and the admission mobility score.

S.6. Denominator Statement: The target population (denominator) for this quality measure is the number of LTCH patients requiring ventilator support at the time of admission to the LTCH.

S.8. Denominator Exclusions: This quality measure has following patient-level exclusion criteria:

1) Patients with incomplete stays:

Rationale: When a patient has an incomplete stay, for example, the patients leave urgently due to a medical emergency, it can be challenging to gather accurate discharge functional status data. Patients with incomplete stays include patients who are unexpectedly discharged to an acute-care setting (Inpatient Prospective Payment System or Inpatient Psychiatric Hospital or unit) because of a medical emergency or psychiatric condition; patients transferred to another LTCH; patients who leave the LTCH against medical advice; patients who die; and patients with a length of stay less than 3 days.

2) Patients discharged to hospice:

Rationale: Patients discharged to hospice are excluded because functional improvement may not be a goal for these patients.

3) Patients with progressive neurological conditions, including amyotrophic lateral sclerosis, multiple sclerosis, Parkinson's disease, and Huntington's chorea:

Rationale: These patients are excluded because they may have functional decline or less predictable function trajectories.

4) Patients in coma, persistent vegetative state, complete tetraplegia, and locked-in syndrome:

Rationale: The patients are excluded because they may have limited or less predictable mobility recovery.

5) Patients younger than age 21:

Rationale: There is only limited evidence published about functional outcomes for individuals younger than 21.

6) Patients who are coded as independent on all the mobility items at admission:

Rationale: These patients are excluded because no improvement in mobility skills can be measured with the mobility items used in this quality measure.

Facility-level quality measure exclusion: For LTCHs with fewer than 20 patient stays, data for this quality measure are not publicly reported.

De.1. Measure Type: Outcome

S.17. Data Source: Instrument-Based Data

S.20. Level of Analysis: Facility

IF Endorsement Maintenance – Original Endorsement Date: Jul 23, 2015 **Most Recent Endorsement Date:** Jul 23, 2015

IF this measure is included in a composite, NQF Composite#/title:

IF this measure is paired/grouped, NQF#/title:

De.4. IF PAIRED/GROUPED, what is the reason this measure must be reported with other measures to appropriately interpret results? Not applicable. This measure is not paired or grouped with another measure.

Preliminary Analysis: Maintenance of Endorsement

To maintain NQF endorsement endorsed measures are evaluated periodically to ensure that the measures still meets the NQF endorsement criteria ("maintenance"). The emphasis for maintaining endorsement is focused on how effective the measure is for promoting improvements in quality. Endorsed measures should have some experience from the field to inform the evaluation. The emphasis for maintaining endorsement is noted for each criterion.

Criteria 1: Importance to Measure and Report

1a. Evidence

Maintenance measures – less emphasis on evidence unless there is new information or change in evidence since the prior evaluation.

1a. Evidence. The evidence requirements for a health outcome measure include providing empirical data that demonstrate a relationship between the outcome and at least one healthcare structure, process, intervention, or service; if these data not available, data demonstrating wide variation in performance, assuming the data are from a robust number of providers and results are not subject to systematic bias. For measures derived from patient report, evidence also should demonstrate that the target population values the measured outcome, process, or structure and finds it meaningful.

Evidence Summary

- Long term care hospitals (LTCH) treat patients who are chronically critically ill; use is increasing over time. Many of these patients are on ventilators and have functional limitations. Research suggests early mobilization/additional physical therapy for these patients is associated with improved outcomes and either improvement in functional status or less functional decline at discharge.

Summary of prior review in 2014

- This outcome measure estimates the risk-adjusted change in mobility score between admission and discharge among LTCH patients requiring ventilator support at admission.
- As such, the evidence should include a rationale that identifies and supports the relationship of the outcome to at least one healthcare structure, process, intervention, or service.
- The developer states that "Functional improvement is particularly relevant for patients who require ventilator support because these patients traditionally have limited or no mobility because of cardiovascular and pulmonary instability, delirium, sedation, lack of rehabilitation therapy staff, and lack of physician referral (Zanni et al., 2010). A Medicare Payment Advisory Commission (2014) analysis of Medicare Provider Analysis and Review data found that 16 percent of LTCH patients used at least one ventilator-related service in 2012."

Changes to evidence from last review

☐ **The developer attests that there have been no changes in the evidence since the measure was last evaluated.**

☒ **The developer provided updated evidence for this measure:**

Updates:

- Developer offered a logic model depicting the relationship between structures, processes and patient mobility outcomes.
- Long-term care hospitals (LTCHs), sometimes referred to as long-term acute care hospitals, treat patients who are chronically critically ill. Use has increased in the last 20 years. In calendar year 2017, there were 414 LTCHs providing care for 161,886 cases. The average length of stay in an LTCH is 27.0 days and the average Medicare cost per case in 2017 was $38,253. Patients have limitations on function and complex medical needs.
- Developers conducted a literature review for new literature since 2014. There is limited literature for LTCHs generally, so they also include several studies that addressed mobility outcomes for patients requiring ventilator support treated in intensive care units, and a clinical practice guideline that recommended early mobilization.
- Found three studies: a randomized pilot trial, an observational cohort study, and a case report.
    - Randomized pilot trial found patients with the added therapy intervention had improved strength, physical function and mobility at discharge compared to usual physical therapy LTCH patients. Further, the added therapy was also associated with greater success being weaned off the ventilator and being discharged home than usual care alone.
    - Observational cohort study found patients receiving therapy demonstrated significant improvements in function between admission and discharge using the Functional Status Scale -Intensive Care Unit.
    - Case report found graded mobilization program using a mobile leg press and a hydraulic-assist platform walker) which the authors suggested expedited the patient's recovery process.
- Other research that included (but did not focus on) functional status for ventilator patients in various types of facilities found physical therapy interventions helpful and early mobilization associated with improved status at discharge.

Question for the Committee:

- *Is there at least one thing that the provider can do to achieve a change in the measure results?*

Questions for the Committee:

- The evidence provided by the developer is updated, directionally the same, and stronger compared to that for the previous NQF review. Does the Committee agree there is no need for repeat discussion and vote on Evidence?

Guidance from the Evidence Algorithm

Measure assesses outcome (box 1) YES -> relationship between outcome and at least one healthcare action (box 2) YES -> PASS

Preliminary rating for evidence: ☒ Pass ☐ No Pass

1b. Gap in Care/Opportunity for Improvement and 1b. Disparities

Maintenance measures – increased emphasis on gap and variation

1b. Performance Gap. The performance gap requirements include demonstrating quality problems and opportunity for improvement.

- Patients in LTCHs are at high risk for functional decline during their stay, and patients on ventilators are at especially high risk of decline in functional status. According to CDC, "more than 300,000 patients receive mechanical ventilation in the United States each year (CDC 2014). These patients are at increased risk for infections, such as pneumonia and sepsis, as well as other serious complications including pulmonary edema, pulmonary embolism, and death (Esteban et al 2002; Klompas et. al 2011). These complications can lead to longer stays in the intensive care unit and hospital, increased health care costs and increased risk of disability (or death) (CDC 2014). The estimated mortality rate in patients aged 85 years and older with acute lung injury on mechanical ventilation is 60 percent (Rubenfeld et al. 2005)."
- 16% of LTCH patients used at least one ventilator-related service in 2012, and this has likely increased.
- Scores have been stable over the last five rolling four quarters: mean scores ranged from 9.0 to 9.2. The interquartile range for the rolling four quarters data ranged from 4.1 to 4.5 mobility units. Across eight quarters (Q3 2016 – Q2 2018), quality measure score distributions showed variation in LTCH outcomes. Between Q3 2016 to Q2 2018, the overall mean decreased marginally from 9.2 to 9.1.

Disparities

- The developers assessed scores for four social risk factors (payor source, marriage status, race, and ethnicity) to assess whether there are disparities in cares. 31.2% of patients were Medicaid beneficiaries, 66.0% of patients were white, and 39.7% were not currently married.
- The unadjusted mean change in mobility score varied by race/ethnicity and payer source.
    - Black (6.6) and Asian (6.3) patients had lower unadjusted change in mobility scores than White (9.5) patients.
    - Medicaid (8.0) patients, who were likely dual-eligible, had lower unadjusted change scores than both patients with Medicare (8.4) and private insurance (9.5).
    - Patients not currently married (8.4) also had lower unadjusted changes scores than patients who are married (9.1).
    - Mean change in mobility scores were similar when examining ethnicity (8.7 for non-Hispanic and 8.5 for Hispanic).
- Further analysis of social risk factors and their relationships to patients' change in mobility scores indicate that some factors (Medicaid payer, Black race) were tied to lower mobility change scores while

others (Hispanic ethnicity, American Indian or Alaska Native race) were tied to higher mobility change scores.

- The developer was unable to identify any literature that reported on disparities related to functional improvement amount LTCH patients.

Question for the Committee:

- Does the performance gap reported by the developer represent opportunity for improvement that can be addressed by the measured entities?

Preliminary rating for opportunity for improvement: ☒ **High** ☐ **Moderate** ☐ **Low** ☐ **Insufficient**

Committee Pre-evaluation Comments:
Criteria 1: Importance to Measure and Report (including 1a, 1b, 1c)

1a. Evidence: *For all measures (structure, process, outcome, patient-reported structure/process), empirical data are required. How does the evidence relate to the specific structure, process, or outcome being measured? Does it apply directly or is it tangential? How does the structure, process, or outcome relate to desired outcomes? For maintenance measures –are you aware of any new studies/information that changes the evidence base for this measure that has not been cited in the submission? For measures derived from a patient report: Measures derived from a patient report must demonstrate that the target population values the measured outcome, process, or structure.*

- No signficant new data
- PASS. No need to discuss

1b. Performance Gap: *Was current performance data on the measure provided? How does it demonstrate a gap in care (variability or overall less than optimal performance) to warrant a national performance measure? Disparities: Was data on the measure by population subgroups provided? How does it demonstrate disparities in the care?*

- Measured for race, ethnicity, marital status and insurance types. There were statistically significant differences between some subgroups and the mean. The measure shows a wide variety of performace differences, however it is unclear how this translates into specific outcomes.
- High opportunity for improvement

Criteria 2: Scientific Acceptability of Measure Properties

2a. Reliability: Specifications **and** Testing

2b. Validity: Testing**;** Exclusions**;** Risk-Adjustment**;** Meaningful Differences**;** Comparability**;** Missing Data

Reliability

2a1. Specifications requires the measure, as specified, to produce consistent (reliable) and credible (valid) results about the quality of care when implemented. For maintenance measures – no change in emphasis – specifications should be evaluated the same as with new measures.

2a2. Reliability testing demonstrates if the measure data elements are repeatable, producing the same results a high proportion of the time when assessed in the same population in the same time period and/or that the measure score is precise enough to distinguish differences in performance across providers. For maintenance measures – less emphasis if no new testing data provided.

Validity

2b2. Validity testing should demonstrate the measure data elements are correct and/or the measure score correctly reflects the quality of care provided, adequately identifying differences in quality. For maintenance measures – less emphasis if no new testing data provided.

2b2-2b6. Potential threats to validity should be assessed/addressed.

Complex measure evaluated by Scientific Methods Panel? ☒ **Yes** ☐ **No**

Evaluators: NQF Scientific Methods Panel

Methods Panel Review (Combined)

Scientific Methods Panel Votes: Measure Passes

- Reliability: H-2, M-4, L-0, I-0
- Validity: H-2, M-3, L-0, I-1

Standing Committee Summary

The NQF Scientific Methods Panel reviewed this measure and elected not to discuss it further based on achieving a consensus that the measure should pass to the Standing Committee.

Reliability

- Testing included score-level and data element testing
- The developer conducted reliability testing for both data element and measure score. For data element reliability, the developer reported internal consistency, inter-rater reliability (this is relevant because the mobility score is assessed by clinicians using the instrument). In addition, the developer reported the results from the video reliability study. For measure score reliability, the developer conducted split-half reliability testing. The Pearson Product-Moment Correlation (r), Spearman Rank Correlation (ρ), and Intraclass Correlation Coefficient (ICC) were used.
- SMP assessed the reliability results as follows: "Cronbach's alpha was high for the mobility data elements (alpha = 0.92). The facility-level reliability estimates were fair for LTCHs in the lowest quartile of discharges (20-44 discharges) and was more acceptable at higher levels of discharges. For the full sample, the reliability coefficient was 0.71, which is acceptable. Inter-rater reliability of the mobility data elements was reasonable to good."
- SMP note to measure developer on specifications:
    - "Only concern is the scoring of mobility items that contained 'activity not attempted' values (07. Patient refused, 09. Not applicable, 10. Not attempted due to environmental limitations, and 88. Not attempted due to medical condition or safety concerns) or were skipped, dashed, or missing are recoded to 01. This could create the appearance of a decrease in mobility when in fact it was due to factors outside the facilities control. It is noted that missing data was minimal but no analysis of other factors that would have resulted in a score of 01." Also, SMP noted that "a substantial proportion of patients will be excluded due to incomplete stays, while the exclusion is understandable, it will be helpful to see if the proportion of exclusion varies significantly across facilities."

Validity

- Testing included score-level and data element testing
- The measure developer assessed construct validity of the mobility data by examining the relation between discharge functional abilities and the discharge destination. Rasch analysis uses item data to determine how well items in a scale/instrument function together to measure a construct. Content validity was evaluated by comparing other mobility measurement instruments. Score validity by a logistic regression model to examine the association between observed discharge mobility scores and the odds of a community discharge.

- SMP members summarized the results as follows: "Content validity results were very good. However, the results for both data elements and scale construct validity were to be expected, not particularly convincing." Hence, a moderate validity rating.
- SMP notes/comments to measure developer:
    - "A very high proportion of patient stays (37.7%) were excluded due to incomplete stays. It would be useful and important to know if the exclusion varies significantly across LTCH facilities."

Standing Committee Action Item:

- The Standing Committee can discuss reliability and/or validity or accept the Scientific Methods Panel ratings.

Questions for the Committee regarding reliability:

- Do you have any concerns that the measure can be consistently implemented (i.e., are measure specifications adequate)?
- The Scientific Methods Panel is satisfied with the reliability testing for the measure. Does the Committee think there is a need to discuss and/or vote on reliability?

Questions for the Committee regarding validity:

- Do you have any concerns regarding the validity of the measure (e.g., exclusions, risk-adjustment approach, etc.)?
- The Scientific Methods Panel is satisfied with the validity analyses for the measure. Does the Committee think there is a need to discuss and/or vote on validity?

Preliminary rating for reliability:	☐ **High**	☒ **Moderate**	☐ **Low**	☐ **Insufficient**
Preliminary rating for validity:	☐ **High**	☒ **Moderate**	☐ **Low**	☐ **Insufficient**

Combined Methods Panel Scientific Acceptability Evaluation

Evaluating Scientific Acceptability: Instructions

Measure Number: 2632

Measure Title: Long-Term Care Hospital (LTCH) Functional Outcome Measure: Change in Mobility Among Patients Requiring Ventilator Support

Type of measure:

☐ **Process** ☐ **Process: Appropriate Use** ☐ **Structure** ☐ **Efficiency** ☐ **Cost/Resource Use**

☒ **Outcome** ☐ **Outcome: PRO-PM** ☐ **Outcome: Intermediate Clinical Outcome** ☐ **Composite**

Data Source:

☒☐ **Claims** ☐ **Electronic Health Data** ☐ **Electronic Health Records** ☐ **Management Data**

☒☐ **Assessment Data** ☐ **Paper Medical Records** ☐☒ **Instrument-Based Data** ☐ **Registry Data**

☐ **Enrollment Data** ☒☐ **Other** LTCH CARE Data set

Level of Analysis:

☐ **Clinician: Group/Practice** ☐ **Clinician: Individual** ☒ **Facility** ☐ **Health Plan**

☐ **Population: Community, County or City** ☐ **Population: Regional and State**

☐ **Integrated Delivery System** ☐ **Other**

Measure is:

☐ **New** ☒ **Previously endorsed (**NOTE: Empirical validity testing is expected at time of maintenance review; if not possible, justification is required.)

Methods Panel Member (MP) 1: This measure estimates the risk-adjusted change in mobility score between admission and discharge among LTCH patients requiring ventilator support at admission The measure does not have a simple form for the numerator and denominator. This measure estimates the risk-adjusted change in mobility score between admission and discharge among LTCH patients requiring ventilator support at admission. The change in mobility score is calculated as the difference between the discharge mobility score and the admission mobility score. The target population (denominator) for this quality measure is the number of LTCH patients requiring ventilator support at the time of admission to the LTCH.

RELIABILITY: SPECIFICATIONS

1. **Are submitted specifications precise, unambiguous, and complete so that they can be consistently implemented?** ☒ **Yes** ☐ **No**

   Submission document: "MIF_xxxx" document, items S.1-S.22

   NOTE*: NQF staff will conduct a separate, more technical, check of eCQM specifications, value sets, logic, and feasibility, so no need to consider these in your evaluation.*

2. **Briefly summarize any concerns about the measure specifications.**

   MP 2: No

   MP 3: Only concern is the scoring of mobility items that contained 'activity not attempted' values (07. Patient refused, 09. Not applicable, 10. Not attempted due to environmental limitations, and 88. Not attempted due to medical condition or safety concerns) or were skipped, dashed, or missing are recoded to 01. This could create the appearance of a decrease in mobility when in fact it was due to factors outside the facilities control. It is noted that missing data was minimal but no analysis of other factors that would have resulted in a score of 01.

   MP 5 MP 5:A substantial proportion of patients will be excluded due to imcomplete stays, while the exclusion is understandable, it will be helpful to see if the proportion of exclusion varies significantly across facilities.

RELIABILITY: TESTING

Submission document: "MIF_xxxx" document for specifications, testing attachment questions 1.1-1.4 and section 2a2

3. **Reliability testing level** ☐☒ **Measure score** ☒ **Data element** ☐ **Neither**
4. **Reliability testing was conducted with the data source and level of analysis indicated for this measure** ☒ **Yes** ☐ **No**
5. If score-level and/or data element reliability testing was NOT conducted or if the methods used were NOT appropriate, was **empirical <u>VALIDITY</u> testing** of <u>patient-level data</u> conducted?

   ☒☐ **Yes** ☐ **No** NA

6. **Assess the method(s) used for reliability testing**

   Submission document: Testing attachment, section 2a2.2

   MP 1:Heavy emphasis on interrater reliability. Also conducted internal consistency (Cronbach's alpha)

   MP 2: Testing methods are appropriate.

   MP 5: The developer conducted reliability testing for both data element and measure score. For data element reliability, the developer reported internal consistency, inter-rater reliability (this is relevant because the mobility score is assessed by clinicains using the instrument). In addition, the developer reported the

results from the video reliability study. For measure score reliability, the developer conducted split-half reliability testing. Both tests were appropriate.

MP 4: For data element reliability, a Cronbach's alpha measure of internal consistent was used – this is a standard method for multi-item scales, although a little unusual when the items are intentionally designed to be scaled in their difficulty. It's not clear how high the Cronbach's alpha results SHOULD be if it expected that a patient at a given level of mobilitywill be able to do some things independently, other things with assistance, and perhaps not be able to do some things at all. For measure score reliability, a split-half approach was used in which the cases for each facility were divided in half, with scores derived from each half, and then the results compared for the two half-samples, using three different statistical tests.

MP 3: Split-half reliability was used to examine the reliability of the computed performance measure scores. The Pearson Product-Moment Correlation (r), Spearman Rank Correlation (ρ), and Intraclass Correlation Coefficient (ICC) were used to examine the performance measure reliability. Intraclass correlations were also calculated by facility volume quartile to examine whether there were differences in performance measure reliability by LTCH size. Internal consistency was assessed using the Cronbach's alpha coefficient. Inter-Rater Reliability based on previous studies

MP 6: The developer used Cronbach's alpha to evaluate internal consistency of the mobility items and used an ICC with a split half approach to evaluate facility level reliability. The developer used risk-adjusted scores in the ICC analysis. The developer also conducted inter-rater reliability of the data elements. All of the approaches were appropriate, although bootstrapping is important to add confidence to the findings for the facility level reliability estimates.

7. **Assess the results of reliability testing**

Submission document: Testing attachment, section 2a2.3

MP 3: Previous IRR demonstrated acceptable Kappa scores.): Split-half analysis results indicated positive moderate-to-strong correlations. ICCs for the volume quartiles showed moderate to strong scores.

MP 1: Internal consistency good (range 0.75-0.98). Interrater reliability (ICCs) good with n>44; adequate (0.60) for n=20-44. Unreliable for facilities with fewer than 20 patients.

Submission document: Testing attachment, section 2a2.3

MP 6: Cronbach's alpha was high for the mobility data elements (alpha = 0.92). The facility-level reliability estimates were fair for LTCHs in the lowest quartile of discharges (20-44 discharges), and was more acceptable at higher levels of discharges. For the full sample, the reliability coefficient was 0.71, which is acceptable. Inter-rater reliability of the mobility data elements was reasonable to good.

MP 2: Test sample is adequate. Moderate to High reliability.

MP 5: Cronbach's alpha was very good, inter-rater reliability measured by weighted kappa ranged from moderate to very high. In general, results for data element reliabity testing were very good. However, one aspect of the video reliability study restuls was concerning, that is, for some items, there were variable levels of agreement across clinical disciplines. Overall agreement was also moderate, ranging from 50% to 78%. This calls into question if it is necessary to account for the potential difference among clinical raters.

Measure score reliabliyt measured by ICC was also acceptable.

MP 5: The results of testing at both data element and measure relability showed acceptable reliability, using standard and generally-accepted methods.

8. Was the method described and appropriate for assessing the proportion of variability due to real differences among measured entities? NOTE: If multiple methods used, at least one must be appropriate.

Submission document: Testing attachment, section 2a2.2

☒ **Yes**

☐ **No**

☐ **Not applicable** (score-level testing was not performed)

9. Was the method described and appropriate for assessing the reliability of ALL critical data elements?

   Submission document: Testing attachment, section 2a2.2

   ☒ **Yes**

   ☐ **No**

   ☐ **Not applicable** (data element testing was not performed)

10. **OVERALL RATING OF RELIABILITY** (taking into account precision of specifications and all testing results):

    ☐☒ **High** (NOTE: Can be HIGH only if score-level testing has been conducted)

    ☒☐ **Moderate** (NOTE: Moderate is the highest eligible rating if score-level testing has not been conducted)

    ☐ **Low** (NOTE: Should rate LOW if you believe specifications are NOT precise, unambiguous, and complete or if testing methods/results are not adequate)

    ☐ **Insufficient** (NOTE: Should rate INSUFFICIENT if you believe you do not have the information you need to make a rating decision)

11. **Briefly explain rationale for the rating of OVERALL RATING OF RELIABILITY and any concerns you may have with the approach to demonstrating reliability.**

    MP 2: Based on testing of the sample.

    MP 3: All analysis demonstrated reliability. No concerns noted

MP 5: Comprehensive reliability testing was conducted, covering both data element and measure score. The results were in general good with the exception of somewhat variable agreement on some items across clinical disciplines.

MP 5: As noted above, the results of reliability tests were generally positive, and the measure score reliability depends (as it normally does) on having an adequate sample size.

MP 6: Bootstrapping the sample would make the facility testing results more robust.

VALIDITY: ASSESSMENT OF THREATS TO VALIDITY

12. **Please describe any concerns you have with measure exclusions.**

    Submission document: Testing attachment, section 2b2.

    MP 1: 299 of 750 patients excluded! I don't understand the following exclusions, as it seems they could be included and handled with risk adjustment or other means. If rates of these differ, it could throw an advantage to those hospitals with more such cases. If they do not differ, then it is not an issue to include:

    - Patients with incomplete stays
    - Patients discharged to hospice

    Patients with progressive neurological conditions, including amyotrophic lateral sclerosis, multiple sclerosis, Parkinson's disease, and Huntington's chorea

    MP 2: No concerns

    MP 5:A very high proportion of patient stays (37.7%) were exclulded due to incomplete stays. It would be useful and important to know if the exclusion varies signficiantly across LTCH facilities.

    MP 3: No concerns

    MP 6: None –; exclusions appear appropriate.

    MP 5: None – exclusions seemed appropriate

13. **Please describe any concerns you have regarding the ability to identify meaningful differences in performance.**

Submission document: Testing attachment, section 2b4.

MP 1: Extensive missing cases risks an unrepresentative sample

MP 2: None

MP 5: No concern.

MP 5: Although the developers were able to show that a number of facilities had performance that was significantly above or below a national average, it is not clear whether those differences are clinically meaningful to patients or family members.

MP 3: None

MP 6: None.

14. **Please describe any concerns you have regarding comparability of results if multiple data sources or methods are specified.**

Submission document: Testing attachment, section 2b5.
MP 2:NA
MP 5: No concern.
MP 3: None
MP 6:N/A

15. **Please describe any concerns you have regarding missing data.**

Submission document: Testing attachment, section 2b6.

MP 2: None

MP 5: No concern

MP 3: As noted previously, missing scores calculated to 01

MP 6: None.

16. **Risk Adjustment**

16a. **Risk-adjustment method** ☐ **None** ☒ **Statistical model** ☐ **Stratification**

16b. **If not risk-adjusted, is this supported by either a conceptual rationale or empirical analyses?**

☒☐ Yes ☐ No ☐☒ Not applicable

16c. **Social risk adjustment:**

MP 1:4 social risk factors affected performance measure scores. The social risk factors are: 1) payer source (patient-level variable); 2) marriage status (patient-level variable); 3) race (patient-level variable); and 4) ethnicity (patient-level variable).

16c.1 Are social risk factors included in risk model? ☐☒ Yes ☒☐ No ☐ Not applicable

16c.2 Conceptual rationale for social risk factors included? ☐☒ Yes ☒☐ No

16c.3 Is there a conceptual relationship between potential social risk factor variables and the measure focus? ☐☒ Yes ☒☐ No

16d.**Risk adjustment summary:**

16d.1 All of the risk-adjustment variables present at the start of care? ☒ Yes ☐ No

16d.2 If factors not present at the start of care, do you agree with the rationale provided for inclusion? ☐ Yes ☐ No **MP 2:**NA

16d.3 Is the risk adjustment approach appropriately developed and assessed? ☐☒ Yes ☐ No

16d.4 Do analyses indicate acceptable results (e.g., acceptable discrimination and calibration) ☒ Yes ☐ No

16d.5.Appropriate risk-adjustment strategy included in the measure? ☒ Yes ☐ No

16e**. Assess the risk-adjustment approach**

MP 1:22 risk factors included Recently updated the risk adjustment model by removing several comorbidities. They found some comorbidities were no longer significant predictors of change in mobility or the association between the comorbidity and functional outcomes was no longer consistent with the evidence from the literature or clinical expectations.

The risk-adjustment procedure involves comparing patients' observed change in mobility scores with their expected change in mobility scores. The prior approach used the ratio of the observed to expected values and the ratio was multiplied by the national mean. The new approach uses the difference between the observed and expected values, and the difference value is added to the national mean. This seems a good change.

MP 5: Overall, the risk adjustment approach was acceptable.

MP 2: Appropriate methods used.

MP 5: The approach was generally thoughtful and acceptable, but the developers found that some racial or ethnic groups had significantly different outcomes, and those factors could have been included in the adjustment modelts, but the developers fell back on the standard CMS "more research is necessary to understand this phenomenon" rationale for not including these social factors. Unfortunately, a decision to either include or exclude the factors has consequences for the affected providers, for patients and families, and for other stakeholders, so a decision to not include social factors like race or ethnicity because of "insufficient research" is itself a decision with potential adverse consequences for facilities serving minority patients, whose performance will appear to be worse than it actually may be. The data on the small effects of risk adjustment isn't entirely convincing, as one would expect the mean and median scores in a distribution to not change with adjustment; the key issue is how many individual facilities would move up or down by some defined amount in the distribution with adjustment. It may be true that no facilities would have moved much in the distribution with a more-inclusive adjustment model, but that information doesn't seem to have been provided.

MP 3: Agree with submitors that social risk facors did not impact the results so no adjustment for risk factors needed

MP 6: The risk adjustment approach is completely empirically derived and accounts for a reasonable amount of variance in change in mobility scores.

<u>For cost/resource use measures ONLY</u>:

17. **Are the specifications in alignment with the stated measure intent?**

    x Yes ☐ Somewhat ☐ No (If "Somewhat" or "No", please explain)

18. **Describe any concerns of threats to validity related to attribution, the costing approach, carve outs, or truncation (approach to outliers):**

VALIDITY: TESTING

19. **Validity testing level: ☐ Measure score ☒☐ Data element ☐☒ Both**

20. **Method of establishing validity of the measure score:**

    ☒☐ Face validity

    ☐☒ Empirical validity testing of the measure score

    ☒☐ N/A (score-level testing not conducted)

21. **Assess the method(s) for establishing validity**

    Submission document: Testing attachment, section 2b2.2

MP 3:construct validity of the mobility data by examining the relation between discharge functional abilities and the discharge destination. Score validity by a logistic regression model to examine the association between observed discharge mobility scores and the odds of a community discharge. Rasch analysis uses item data to determine how well items in a scale/instrument function together to measure a construct. Content validity by comparing other mobility measurement instruments.

MP 1:

Extensive use of Rasch Measurement to evaluate model fit and item fit, supporting sufficient unidimensionality of mobility.

1. **Data element Construct Validity – Observed Discharge Mobility Scores by Discharge Destination (unit of analysis is patient stays):**
2. **Scale/Instrument Construct Validity – Observed Discharge Mobility Scores and Discharge Destination (unit of analysis is patient stays):**
3. **Scale/Instrument Construct Validity – Data Element (Item) Difficulty Ordering Using Rasch Analysis (unit of analysis is patient assessment data):**
4. **Data Element (Item) and Scale/Instrument Validity - Fit Assessment Analysis (unit of analysis is patient assessment data):**
5. **Data Element (Item) and Scale/Instrument Validity - Response Option Assessment Using Rasch Analysis (unit of analysis is patient assessment data):**

MP 2: Appropriate

MP 5: Extensive data element validity tests were conducted, including content validity, data elements constrcut validity, scale construct validity, and others. For both data elements and scale construct validity testing, the developer correlated discharge destination with both the data elements and scale and considered positive relationship as evidence of validity. It can be argued that discharge destination is not an ideal evaluation criterion as discharge destination may be partly determined by the mobiliyt score assessed.

MP 5: A number of tests of different types of validity were done on the mobility instrument, and all were appropriate and consistent with commonly-used methods, including the use of Rasch analysis to establish the differential difficulty of items in the mobility scale.

MP 6: The developer used several approaches to establish validity, ranging from an analysis and comparison of the data elements with other mobility assessments to a logistic regression model examining the association between mobility scores and discharge to the community. The authors also conducted a number of rasch analyses to determine construct fit. The most compelling approach was the logit model.

22. **Assess the results(s) for establishing validity**

Submission document: Testing attachment, section 2b2.3

MP 2: Test sample is adequate. Sufficient validity

MP 5: Content validity results were very good. However, the results for both data elements and scale construct validity were to be expected, not particularly convincing.

MP 5: Validity at the data element level (mobility assessment instrument) was good. No validity testing at the score level was done, even though the developers checked that box on the testing form

MP 3: Mobility data elements data were positively associated with discharge destination

MP 6: Overall, the developer successfully demonstrated the validity of this measure through the use of several approaches, the most compelling of which were the regression analysis and rasch analyses.

23. **Was the method described and appropriate for assessing conceptually and theoretically sound hypothesized relationships?**

Submission document: Testing attachment, section 2b1.

☒ **Yes** ☐☒ **Yes , MP 6:**with some concerns noted below

☐ **No**

☒☐ **Not applicable** (score-level testing was not performed)

24. **Was the method described and appropriate for assessing the accuracy of ALL critical data elements?** *NOTE that data element validation from the literature is acceptable.*

    Submission document*: Testing attachment, section 2b1.*

    ☒ **Yes**

    ☐ **No**

    ☐ **Not applicable** (data element testing was not performed)

25. **OVERALL RATING OF VALIDITY taking into account the results and scope of all testing and analysis of potential threats.**

    ☐☒ **High** (NOTE: Can be HIGH only if score-level testing has been conducted)

    ☒☐ **Moderate** (NOTE: Moderate is the highest eligible rating if score-level testing has NOT been conducted)

    ☐ **Low** (NOTE: Should rate LOW if you believe that there are threats to validity and/or relevant threats to validity were not assessed OR if testing methods/results are not adequate)

    ☒☐ **Insufficient** (NOTE: For instrument-based measures and some composite measures, testing at both the score level and the data element level is required; if not conducted, should rate as INSUFFICIENT.)

26. **Briefly explain rationale for rating of OVERALL RATING OF VALIDITY and any concerns you may have with the developers' approach to demonstrating validity.**

    MP 2: Based on testing results and comparision with known instruments.

    MP 3: All analysis demonstrated adequate validity. No concerns noted

MP 5: Based on NQF guideline, for instrument-based measure, validity testing at the measure score level is required. This form doesn't provide measure score level validity testing results.

MP 5: Since no measure score validity testing was done, the validity of the measure as a quality of care measure for LTCHs really depends on face validity as assessed by members of the standing committee who will be evaluating this measure. There is no statistical basis for declaring the measure to be a valid measure of quality of care. Data element validity, though, is strong.

MP 6: The logit model provided evidence of a link between discharge scale scores and discharge to the community (i.e., criterion validity). However, there is a likelihood that unmeasured factors in this logit model account for the relationship such as social support and the availability of community based services. At the data element level, validity analyses based on the rasch analyses demonstrate that the items in the scale are valid.

FOR COMPOSITE MEASURES ONLY: Empirical analyses to support composite construction

27. **What is the level of certainty or confidence that the empirical analysis demonstrates that the component measures add value to the composite and that the aggregation and weighting rules are consistent with the quality construct?**

    ☐ **High**

    ☐ **Moderate**

    ☐ **Low**

    ☐ **Insufficient**

28. **Briefly explain rationale for rating of EMPIRICAL ANALYSES TO SUPPORT COMPOSITE CONSTRUCTION**

ADDITIONAL RECOMMENDATIONS

29. **If you have listed any concerns in this form, do you believe these concerns warrant further discussion by the multi-stakeholder Standing Committee? If so, please list those concerns below.**

MP 5: Rasch analysis results indicated that 15 mobility items are of different degrees of difficulty. This implies that the difference between 10 and 20 may not be the same as the differene between 30 and 40. This has implications for across IRFs comparison because this is a measure based on change score.

MP 5: Since the developers provided no evidence of measure score-level validity, the standing committee should be thoughtful in making its own decision about the face validity of this measure as a measure of quality of care for LTCHs.

Committee Pre-evaluation Comments:
Criteria 2: Scientific Acceptability of Measure Properties (including all 2a, 2b, and 2c)

2a1. Specifications: *Which data elements, if any, are not clearly defined? Which codes with descriptors, if any, are not provided? Which steps, if any, in the logic or calculation algorithm or other specifications (e.g., risk/case-mix adjustment, survey/sampling instructions) are not clear? What concerns do you have about the likelihood that this measure can be consistently implemented?*

- None. Inter-rater reliability is good
- Reliability specifications are reasonable

2a2. Reliability testing: *Do you have any concerns about the reliability of the measure?*

- No
- MODERATE reliability

2b2. Validity testing: *Do you have any concerns with the testing results?*

- Yes, the way in which exclusions for incomplete data are handled, approx 40% of cases. Exclusions account for about 50% of all discharges, a very high percentage.
- Moderate validity

Validity- Threats to Validity: *Threats to Validity (Statistically Significant Differences, Multiple Data Sources, Missing Data). 2b4. Meaningful Differences: How do analyses indicate this measure identifies meaningful differences about quality? 2b5. Comparability of performance scores: If multiple sets of specifications: Do analyses indicate they produce comparable results? 2b6. Missing data/no response: Does missing data constitute a threat to the validity of this measure?*

- 2b4-Yes. Unclear how what is measured connects with clinically significant outcomes. Most of the evidence is tangential or from very small studies that do not elaborate on effective clinical interventions. This measure makes sense and has a certain amount of face validity but no compelling data that links specific interventions to outcomes. How should LTACs respond to improving this measure? What are they doing differently because of this measure? It measures a difference but it is not clear that the difference leads to an improvement. Perhaps the release of 2018 data will demonstrate a change. 2b5- ok 2b6- missing data invalidated 37% of results. A very high percentage.
- No threats seen

Other Threats to Validity: *Other Threats to Validity (Exclusions, Risk Adjustment). 2b2. Exclusions: Are the exclusions consistent with the evidence? Are any patients or patient groups inappropriately excluded from the measure? 2b3. Risk Adjustment: If outcome (intermediate, health, or PRO-based) or resource use performance measure: Is there a conceptual relationship between potential social risk factor variables and the measure focus? How well do social risk factor variables that were available and analyzed align with the conceptual description provided? Are all of the risk-adjustment variables present at the start of care (if not, do you agree with the rationale provided)? Was the risk adjustment (case-mix adjustment) appropriately developed and tested? Do analyses indicate acceptable results? Is an appropriate risk-adjustment strategy included in the measure?*

- 2b2- reasonable exclusions which account for approx 10% of cases, all present at admission. 2b3- reasonable approach based on national data set. The social risk factors are reasonable and give results that are different when present but don't make sense clinically. The overall risk adjustment appears reasonable, accounting for approx 20% of the variation. To quote MP 5: "Since no measure score validity testing was done, the validity of the measure as a quality of care measure for LTCHs really depends on face validity as assessed by members of the standing committee who will be evaluating this measure. There is no statistical basis for declaring the measure to be a valid measure of quality of care. Data element validity, though, is strong. Since the developers provided no evidence of measure score-level validity, the standing committee should be thoughtful in making its own decision about the face validity of this measure as a measure of quality of care for LTCHs."
- No threats seen

Criterion 3. Feasibility

Maintenance measures – no change in emphasis – implementation issues may be more prominent

3. Feasibility is the extent to which the specifications including measure logic, require data that are readily available or could be captured without undue burden and can be implemented for performance measurement.

- Generated during provision of care; all data elements in defined fields in electronic health records.
- Software is free and trainings were provided; no costs associated with fees, licensing, etc.

Question for the Committee:

- Does the Committee agree with the staff assessment that there are no significant feasibility challenges associated with this measure?

Preliminary rating for feasibility: ☒ **High** ☐ **Moderate** ☐ **Low** ☐ **Insufficient**

Committee Pre-evaluation Comments:
Criteria 3: Feasibility

3. Feasibility: *Which of the required data elements are not routinely generated and used during care delivery? Which of the required data elements are not available in electronic form (e.g., EHR or other electronic sources)? What are your concerns about how the data collection strategy can be put into operational use?*

- Pass
- High Feasibility

Criterion 4: Usability and Use

Maintenance measures – increased emphasis – much greater focus on measure use and usefulness, including both impact/improvement and unintended consequences

4a. Use (4a1. Accountability and Transparency; 4a2. Feedback on measure)

4a. Use evaluate the extent to which audiences (e.g., consumers, purchasers, providers, policymakers) use or could use performance results for both accountability and performance improvement activities.

4a.1. Accountability and Transparency. Performance results are used in at least one accountability application within three years after initial endorsement and are publicly reported within six years after initial endorsement (or the data on performance results are available). If not in use at the time of initial endorsement, then a credible plan for implementation within the specified timeframes is provided.

Current uses of the measure

Publicly reported? ☒ **Yes** ☐ **No**

Current use in an accountability program? ☒ **Yes** ☐ **No** ☐ **UNCLEAR**

OR

Planned use in an accountability program? ☒ **Yes** ☐ **No**

Accountability program details

- Currently used for quality improvement within Long-Term Care Hospital Quality Reported Program (LTCH QRP)
- 2019 data will be publicly reported on LTCH Compare in 2020 for the LTCH Quality Reporting Program (LTCH QRP)

4a.2. Feedback on the measure by those being measured or others. Three criteria demonstrate feedback: 1) those being measured have been given performance results or data, as well as assistance with interpreting the measure results and data; 2) those being measured and other users have been given an opportunity to provide feedback on the measure performance or implementation; 3) this feedback has been considered when changes are incorporated into the measure

Feedback on the measure by those being measured or others

- Providers receive results and assistance with interpretation via confidential feedback reports, provider training seminars, manuals and materials, and responses to questions submitted to the LTCH QRP Help Desk and LTCH Public Reporting Help Desk.
- Patients and families and other stakeholders can review results on the publicly available LTCH Compare.
- In the 2015 rule proposal, public commenters mostly supported the addition of this measure to the LTCH QRP. The developer also gathered feedback from a TEP in 2017, and some members expressed support. No feedback suggesting changes to the measures was received.

Additional Feedback: N/A

Questions for the Committee:

- How have (or can) the performance results be used to further the goal of high-quality, efficient healthcare?
- How has the measure been vetted in real-world settings by those being measured or others?

Preliminary rating for Use: ☒ **Pass** ☐ **No Pass**

4b. Usability (4a1. Improvement; 4a2. Benefits of measure)

4b. Usability evaluate the extent to which audiences (e.g., consumers, purchasers, providers, policymakers) use or could use performance results for both accountability and performance improvement activities.

4b.1 Improvement. Progress toward achieving the goal of high-quality, efficient healthcare for individuals or populations is demonstrated.

Improvement results

- Trends over time is not yet available; the measure was implemented on April 1, 2016 and requires 8 quarters/24 months of data. 2018-2019 data will be available in the fall of 2020. Preliminary data (4 quarters) was provided in the importance section.

4b2. Benefits vs. harms. Benefits of the performance measure in facilitating progress toward achieving high-quality, efficient healthcare for individuals or populations outweigh evidence of unintended negative consequences to individuals or populations (if such evidence exists).

Unexpected findings (positive or negative) during implementation

- No unexpected findings

Potential harms

- None found

Additional Feedback: N/A

Questions for the Committee:

- How can the performance results be used to further the goal of high-quality, efficient healthcare?
- Do the benefits of the measure outweigh any potential unintended consequences?

Preliminary rating for Usability and use: ☐ **High** ☒ **Moderate** ☐ **Low** ☐ **Insufficient**

Committee Pre-evaluation Comments:
Criteria 4: Usability and Use

4a. Use: *4a1. Use - Accountability and Transparency: How is the measure being publicly reported? Are the performance results disclosed and available outside of the organizations or practices whose performance is measured? For maintenance measures - which accountability applications is the measure being used for? For new measures - if not in use at the time of initial endorsement, is a credible plan for implementation provided? 4a2. Use - Feedback on the measure: Have those being measured been given performance results or data, as well as assistance with interpreting the measure results and data? Have those being measured or other users been given an opportunity to provide feedback on the measure performance or implementation? Has this feedback has been considered when changes are incorporated into the measure?*

- Yes.
- Use PASS

4b. Usability: *4b1. Usability – Improvement: How can the performance results be used to further the goal of high-quality, efficient healthcare? If not in use for performance improvement at the time of initial endorsement, is a credible rationale provided that describes how the performance results could be used to further the goal of high-quality, efficient healthcare for individuals or populations? 4b2. Usability – Benefits vs. harms: Describe any actual unintended consequences and note how you think the benefits of the measure outweigh them.*

- 4b1- insufficient. There needs to be another layer of data linking the functional mobility score to specific outcomes after discharge (e.g. death, hospitalization, SNF use, time home, return to work.) Also need to tease out what specific elements in the interventions that result in improvement are the likely agents. It is unclear how to improve this measure because in many ways it is an "intermediate measure" rather than an outcome measure.
- Moderate usability. No trend data available....yet

Criterion 5: Related and Competing Measures

Related or competing measures

This measure is related to a number of measures:

- 0167 : Improvement in Ambulation/locomotion
- 0175 : Improvement in bed transferring
- 0422 : Functional status change for patients with Knee impairments
- 0423 : Functional status change for patients with Hip impairments
- 0424 : Functional status change for patients with Foot and Ankle impairments
- 0425 : Functional status change for patients with lumbar impairments

- 0428 : Functional status change for patients with General orthopaedic impairments
- 0429 : Change in Basic Mobility as Measured by the AM-PAC:
- 0688 : Percent of Residents Whose Need for Help with Activities of Daily Living Has Increased (long stay)
- 2287 : Functional Change: Change in Motor Score
- 2321 : Functional Change: Change in Mobility Score
- 2612 : CARE: Improvement in Mobility
- 2634 : Inpatient Rehabilitation Facility (IRF) Functional Outcome Measure: Change in Mobility Score for Medical Rehabilitation Patients
- 2636: Inpatient Rehabilitation Facility (IRF) Functional Outcome Measure: Discharge Mobility Score for Medical Rehabilitation Patients
- 2643: Average change in functional status following lumbar spine fusion surgery
- 2653: Average change in functional status following total knee replacement surgery
- 2774: Functional Change: Change in Mobility Score for Skilled Nursing Facilities
- 2775: Functional Change: Change in Motor Score for Skilled Nursing Facilities
- 2776: Functional Change: Change in Motor Score in Long Term Acute Care Facilities
- 2778: Functional Change: Change in Mobility Score for Long Term Acute Care Facilities

Harmonization

- Are the measures harmonized to the extent possible?

Committee Pre-evaluation Comments: Criterion 5: Related and Competing Measures

Related and Competing: *Are there any related and competing measures? If so, are any specifications that are not harmonized? Are there any additional steps needed for the measures to be harmonized?*

- good attempts to harmonize some related SNF measures by changing how the adjusted score is calculated so this model can be used in multiple sites.

Public and Member Comments

Comments and Member Support/Non-Support Submitted as of: June/13/2019

- **No NQF members have submitted support/non-support choices as of this date**

1. Evidence and Performance Gap – Importance to Measure and Report

Extent to which the specific measure focus is evidence-based, important to making significant gains in healthcare quality, and improving health outcomes for a specific high-priority (high-impact) aspect of healthcare where there is variation in or overall less-than-optimal performance. ***Measures must be judged to meet all sub criteria to pass this criterion and be evaluated against the remaining criteria.***

1a. Evidence to Support the Measure Focus – See attached Evidence Submission Form

2632_NQF_evidence_4-22-19.docx

1a.1 For Maintenance of Endorsement: Is there new evidence about the measure since the last update/submission?

Do not remove any existing information. If there have been any changes to evidence, the Committee will consider the new evidence. Please use the most current version of the evidence attachment (v7.1). Please use red font to indicate updated evidence.

Yes

1a. Evidence (subcriterion 1a)

Measure Number (*if previously endorsed*)**:** NQF #2632

Measure Title: Long-Term Care Hospitals (LTCHs) Functional Outcome Measure: Change in Mobility Among Patients Requiring Ventilator Support

IF the measure is a component in a composite performance measure, provide the title of the Composite Measure here: Not applicable

Date of Submission: 4/9/2019

1a.1.This is a measure of: (*should be consistent with type of measure entered in De.1*)

Outcome

☒ Outcome: Change in function: mobility

☐ Patient-reported outcome (PRO):

PROs include HRQoL/functional status, symptom/symptom burden, experience with care, health-related behaviors. (*A PRO-based performance measure is not a survey instrument. Data may be collected using a survey instrument to construct a PRO measure.)*

☐ Intermediate clinical outcome (*e.g., lab value*):

☐ Process:

☐ Appropriate use measure:

☐ Structure:

☐ Composite:

1a.2 LOGIC MODEL Diagram or briefly describe the steps between the healthcare structures and processes (e.g., interventions, or services) and the patient's health outcome(s). The relationships in the diagram should be easily understood by general, non-technical audiences. Indicate the structure, process or outcome being measured.

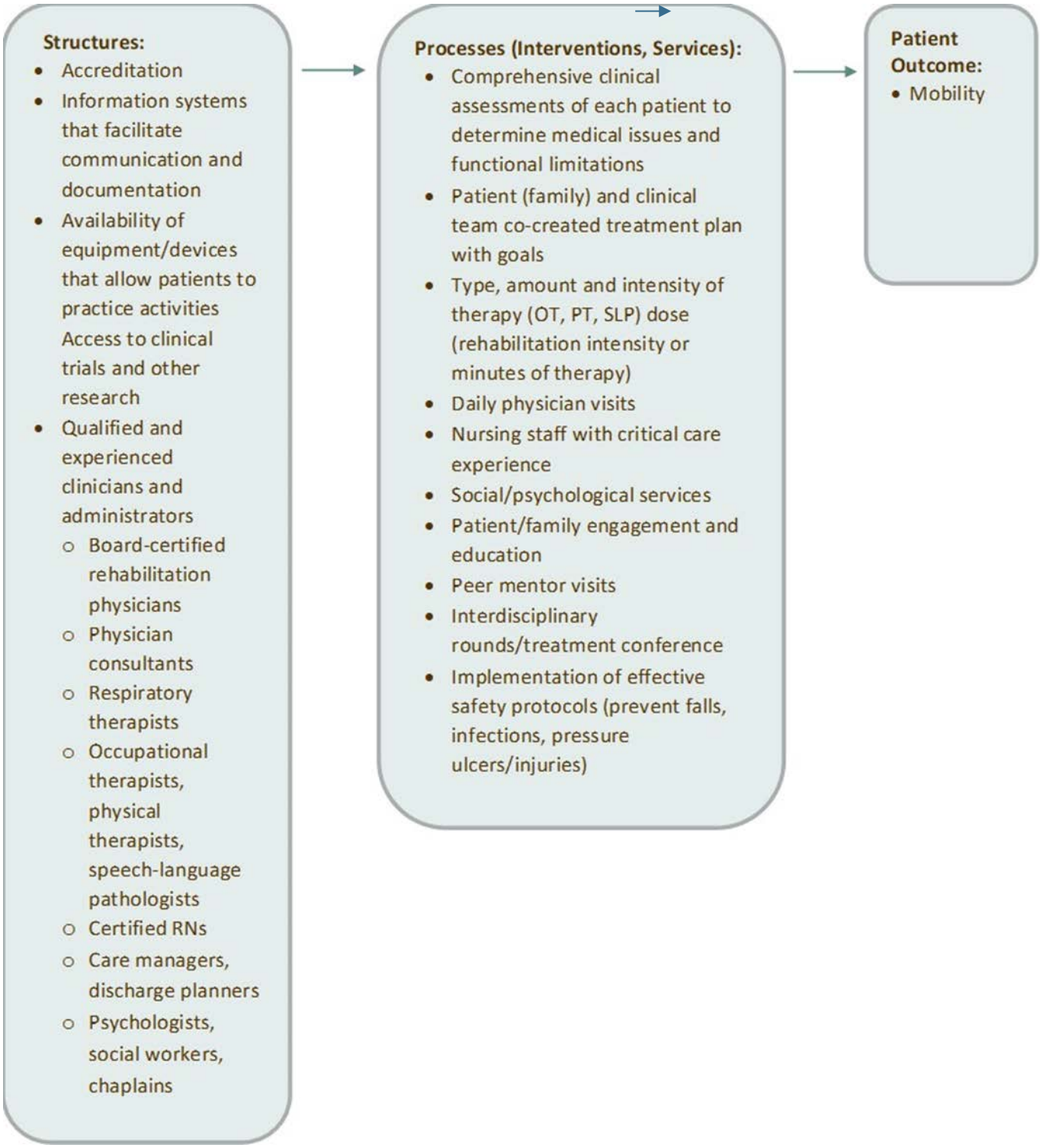
Long-term care hospitals (LTCHs), sometimes referred to as long-term acute care hospitals, treat patients who are chronically critically ill, including those who develop persistent respiratory failure requiring prolonged

mechanical ventilation. Utilization of LTCHs has increased in the last 20 years owing to the increased survival of patients following a critical illness and injury, the aging population, and acute care reimbursement models that incentivize shorter acute care stays. In calendar year 2017, there were 414 LTCHs providing care for 161,886 cases. The average length of stay in an LTCH is 27.0 days and the average Medicare cost per case in 2017 was $38,253.

Many LTCH patients have functional limitations and are at high risk for functional decline during the LTCH stay. In addition to having complex medical care needs for an extended period of time, LTCH patients often have limitations in functioning because of the nature of their conditions, as well as deconditioning due to prolonged bed rest and treatment requirements (e.g., ventilator use). These patients are therefore at high risk for functional deterioration that is both condition-related and iatrogenic (i.e., related to medical treatment). LTCH patients on ventilators are at high risk for functional decline during the hospital stay; therefore, this measure estimates the risk-adjusted change in mobility score between admission and discharge among LTCH patients requiring ventilator support at admission. The change in mobility score is calculated as the difference between the discharge mobility score and the admission mobility score.

Figure 1a lists the structures, processes and the outcomes that relate to this measure. This model shows that an LTCH's structures and processes (treatments or interventions) can result in improved patient functioning.

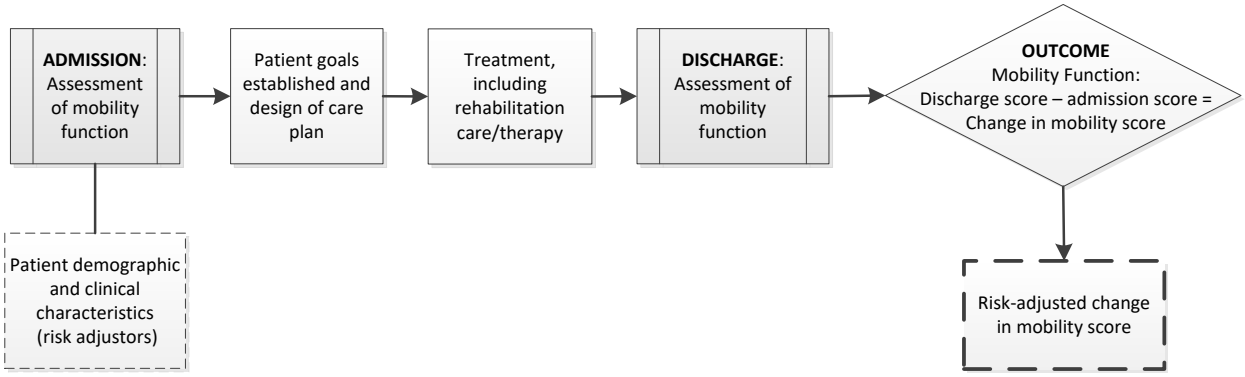
Figure 1a Structures and Processes Associated with Patients' Functional Outcomes.



From 2014 Application:

This outcome measure requires the collection of admission and discharge functional status data by trained clinicians using standardized clinical assessment items (also known as data elements) that assess specific areas of mobility function (see **Figure 1**). The mobility items are coded based on a 6-level rating scale. The Centers for Medicare & Medicaid Services (CMS) intends to revise the LTCH CARE Data Set to include these functional assessment items.

Figure 1. Role of Patient Assessment, Interventions and Functional Outcomes (2014)



1a.3 Value and Meaningfulness: **IF** this measure is derived from patient report, provide evidence that the target population values the measured ***outcome, process, or structure*** and finds it meaningful. (Describe how and from whom their input was obtained.)

****RESPOND TO ONLY ONE SECTION BELOW -EITHER 1a.2, 1a.3 or 1a.4) ****

1a.2 FOR OUTCOME MEASURES including PATIENT REPORTED OUTCOMES - Provide empirical data demonstrating the relationship between the outcome (or PRO) to at least one healthcare structure, process, intervention, or service.

To demonstrate that LTCHs have the ability to improve patient functioning in the area of mobility, NQF requires evidence that at least one structure, process, intervention or service can affect patient functioning. Because therapy services are targeted to improve functional outcomes, we provide a summary of evidence from the literature that is focused on therapy services and functional outcomes.

For this evidence update, we conducted a review of the literature published since our initial NQF application in 2014 to identify relevant manuscripts focused on functional outcomes for patients requiring ventilator support treated in LTCHs. Because there is limited literature focused on LTCHs generally, and few studies that focus on mobility outcomes for LTCH patients requiring ventilator support, we also include several studies that addressed mobility outcomes for patients requiring ventilator support treated in intensive care units, and a clinical practice guideline that recommended early mobilization.

Our literature review found 3 manuscripts that focused on functional outcomes of LTCH patients requiring ventilator support at the time of admission. One study is a pilot trial, one study is an observational cohort study, and one is a case report. Verceles et al.'s (2018) conducted a randomized pilot trial that added a progressive multimodal rehabilitation therapy program to usual care for LTCH patients with ICU-acquired weakness requiring ventilator support. They observed that patients with the added therapy intervention had improved strength, physical function and mobility at discharge compared to usual physical therapy LTCH patients. Further, the added therapy was also associated with greater success being weaned off the ventilator and being discharged home than usual care alone. Thrush et al.'s (2012) observational study examined functional outcomes among LTCH patients, some who were dependent on a ventilator at the time of LTCH admission, and found that patients receiving therapy demonstrated significant improvements in function between admission and discharge using the Functional Status Scale -Intensive Care Unit. A case report (Trees et al., 2013) described novel mobility strategies for managing a patient with ICU-acquired weakness (graded

mobilization program using a mobile leg press and a hydraulic-assist platform walker) which the authors suggested expedited the patient's recovery process.

Many studies of LTCH patients requiring ventilator support have primary and secondary study end points of ventilator liberation and discharge destination and few include therapy services or functional outcomes. We identified two studies that addressed functional outcomes in addition to the primary end points of ventilator liberation and discharge destination. Jubran et al. (2019) described the standard exercise-training program for increasing muscle strength and endurance for patients on ventilators and noted improvements in patient functioning between admission and discharge using Katz summary scores. Sansone et al. (2017) did not describe specific therapy interventions but reported ventilator patients had mean Zubrod scores of 3.0 (± 1.0) at discharge indicating the average patient was spending more than half of the time confined to a bed at the time of discharge.

Our review of recent evidence for patients who require ventilator support treated in intensive care units (ICUs) includes one clinical practice guideline, a quality improvement project, a feasibility study and a multi-site study. Girard et al. (2017), published an Official American Thoracic Society/American College of Chest Physicians Clinical Practice Guideline: Liberation from Mechanical Ventilation in Critically Ill Adults, which included recommendations for protocolized rehabilitation directed toward early mobilization.

Additionally, three studies observed improvement in functional outcomes related to early mobilization in ICU patients requiring mechanical ventilation support. A quality improvement project that promoted early and enhanced rehabilitation within an ICU in the United Kingdom improved patients' mobility status at critical care discharge, and this improvement was associated with reduced ICU and hospital length of stay and reduced days of mechanical ventilation (McWilliams et al, 2015). In a follow-up study, the same team began a randomized trial and published their findings demonstrating the feasibility of introducing a structured program of rehabilitation for patients admitted to critical care. Lastly, a multi-center study (TEAM study investigators, 2015) conducted in Australia and New Zealand found that early mobilization of patients undergoing mechanical ventilation support in the ICU had higher scores in muscle strength (Medical Research Council Manual Muscle Test Sum Score) at discharge than patients not mobilized early. The authors also noticed the large proportion of patients not mobilized early and recommended that early mobilization be a priority in care for patients requiring ventilator support. Clarissa et al., (2019) noted the absence of a standard definition of early mobilization, and highlighted the need for an agreed definition of what constitutes early mobilization in mechanically ventilated patients in order to advance research and practice.

In summary, we described literature that linked rehabilitation therapy interventions and functional outcomes, including mobility outcomes, for LTCH patients requiring ventilator support. Because there are few studies of LTCH patients requiring ventilator support, we supplemented the review with recent studies conducted in ICUs focused on improving functional outcomes of ICU patients who require ventilator support. We believe this literature as a whole demonstrates a link between therapy services and improvement in functioning among patients admitted to the LTCH on a ventilator.

References with Abstracts:

Clarissa C, Salisbury L, Rodgers S, Kean S. Early mobilisation in mechanically ventilated patients: a systematic integrative review of definitions and activities. J Intensive Care. 2019 Jan 17;7:3. doi: 10.1186/s40560-018-0355-z. eCollection 2019.

BACKGROUND: Mechanically ventilated patients often develop muscle weakness post-intensive care admission. Current evidence suggests that early mobilisation of these patients can be an effective intervention in improving their outcomes. However, what constitutes early mobilisation in mechanically ventilated patients (EM-MV) remains unclear. We aimed to systematically explore the definitions and activity types of EM-MV in the literature.

METHODS: Whittemore and Knafl's framework guided this review. CINAHL, MEDLINE, EMBASE, PsycINFO, ASSIA, and Cochrane Library were searched to capture studies from 2000 to 2018, combined with hand search of grey literature and reference lists of included studies. The Critical Appraisal Skills Programme tools were used to

assess the methodological quality of included studies. Data extraction and quality assessment of studies were performed independently by each reviewer before coming together in sub-groups for discussion and agreement. An inductive and data-driven thematic analysis was undertaken on verbatim extracts of EM-MV definitions and activities in included studies.

RESULTS: Seventy-six studies were included from which four major themes were inferred: (1) non-standardised definition, (2) contextual factors, (3) negotiated process and (4) collaboration between patients and staff. The first theme indicates that EM-MV is either not fully defined in studies or when a definition is provided this is not standardised across studies. The remaining themes reflect the diversity of EM-MV activities which depends on patients' characteristics and ICU settings; the negotiated decision-making process between patients and staff; and their interdependent relationship during the implementation.

CONCLUSIONS: This review highlights the absence of an agreed definition and on what constitutes early mobilisation in mechanically ventilated patients. To advance research and practice an agreed and shared definition is a pre-requisite.

Girard TD, Alhazzani W, Kress JP, Ouellette DR, Schmidt GA, Truwit JD, Burns SM, Epstein SK, Esteban A, Fan E, Ferrer M, Fraser GL, Gong MN, Hough CL, Mehta S, Nanchal R, Patel S, Pawlik AJ, Schweickert WD, Sessler CN, Strøm T, Wilson KC, Morris PE; ATS/CHEST Ad Hoc Committee on Liberation from Mechanical Ventilation in Adults. An Official American Thoracic Society/American College of Chest Physicians Clinical Practice Guideline: Liberation from Mechanical Ventilation in Critically Ill Adults. Rehabilitation Protocols, Ventilator Liberation Protocols, and Cuff Leak Tests. Am J Respir Crit Care Med. 2017 Jan 1;195(1):120-133. doi: 10.1164/rccm.201610-2075ST.

BACKGROUND: Interventions that lead to earlier liberation from mechanical ventilation can improve patient outcomes. This guideline, a collaborative effort between the American Thoracic Society and the American College of Chest Physicians, provides evidence-based recommendations to optimize liberation from mechanical ventilation in critically ill adults.

METHODS: Two methodologists performed evidence syntheses to summarize available evidence relevant to key questions about liberation from mechanical ventilation. The methodologists appraised the certainty in the evidence (i.e., the quality of evidence) using the Grading of Recommendations, Assessment, Development, and Evaluation approach and summarized the results in evidence profiles. The guideline panel then formulated recommendations after considering the balance of desirable consequences (benefits) versus undesirable consequences (burdens, adverse effects, and costs), the certainty in the evidence, and the feasibility and acceptability of various interventions. Recommendations were rated as strong or conditional.

RESULTS: The guideline panel made four conditional recommendations related to rehabilitation protocols, ventilator liberation protocols, and cuff leak tests. The recommendations were for acutely hospitalized adults mechanically ventilated for more than 24 hours to receive protocolized rehabilitation directed toward early mobilization, be managed with a ventilator liberation protocol, be assessed with a cuff leak test if they meet extubation criteria but are deemed high risk for postextubation stridor, and be administered systemic steroids for at least 4 hours before extubation if they fail the cuff leak test.

CONCLUSIONS: The American Thoracic Society/American College of Chest Physicians recommendations are intended to support healthcare professionals in their decisions related to liberating critically ill adults from mechanical ventilation.

Jubran A, Grant BJB, Duffner LA, Collins EG, Lanuza DM, Hoffman LA, Tobin MJ. Long-Term Outcome After Prolonged Mechanical Ventilation: A Long-Term Acute-Care Hospital Study. American Journal of Respiratory and Critical Care Medicine In Press. Published on 09-January-2019 as 10.1164/rccm.201806-1131OC

RATIONALE: Patients managed at a long-term acute-care hospital (LTACH) for weaning from

prolonged mechanical ventilation are at risk for profound muscle weakness and disability.

Objectives: To investigate effects of prolonged ventilation on survival, muscle function and its impact on quality of life at six and twelve months after LTACH discharge.

METHODS: Prospective, longitudinal study conducted in 315 patients being weaned from prolonged ventilation at an LTACH.

MEASUREMENTS AND MAIN RESULTS: At discharge, 53.7% of patients were detached from the ventilator and one-year survival was 66.9%. Upon enrollment, maximum inspiratory pressure (PImax) was 41.3 (95% confidence interval [CI], 39.4-43.2) cm H2O (53.1% predicted), whereas handgrip strength was 16.4 (95% CI, 14.4-18.7) kPa (21.5% predicted). At discharge, PImax did not change whereas handgrip strength increased by 34.8% ($p<0.001$). Between discharge and six months, handgrip strength increased 6.2 times more than did PImax. Between discharge and six months, Katz activities-of-daily-living summary-score improved by 64.4%; improvement in Katz summary-score was related to improvement in handgrip strength (rho -0.51;$p<0.001$). By twelve months, physical-summary score and mental-summary score of 36-item Short-Form Survey returned to pre-illness values. When asked, 84.7% of survivors indicated willingness to undergo mechanical ventilation again.

CONCLUSIONS: Among patients receiving prolonged mechanical ventilation at an LTACH, 53.7% were detached from the ventilator at discharge and one-year survival was 66.9%. Respiratory strength was well maintained whereas peripheral strength was severely impaired throughout hospitalization. Six months after discharge, improvement in muscle function enabled patients to perform daily activities, and 84.7% indicated willingness to undergo mechanical ventilation again.

McWilliams D, Weblin J, Atkins G, Bion J, Williams J, Elliott C, Whitehouse T, Snelson C. Enhancing rehabilitation of mechanically ventilated patients in the intensive care unit: a quality improvement project. J Crit Care. 2015 Feb;30(1):13-8. doi: 10.1016/j.jcrc.2014.09.018. Epub 2014 Oct 2.

PURPOSE: Prolonged periods of mechanical ventilation are associated with significant physical and psychosocial adverse effects. Despite increasing evidence supporting early rehabilitation strategies, uptake and delivery of such interventions in Europe have been variable. The objective of this study was to evaluate the impact of an early and enhanced rehabilitation program for mechanically ventilated patients in a large tertiary referral, mixed-population intensive care unit (ICU).

METHOD: A new supportive rehabilitation team was created within the ICU in April 2012, with a focus on promoting early and enhanced rehabilitation for patients at high risk for prolonged ICU and hospital stays. Baseline data on all patients invasively ventilated for at least 5 days in the previous 12 months (n = 290) were compared with all patients ventilated for at least 5 days in the 12 months after the introduction of the rehabilitation team (n = 292). The main outcome measures were mobility level at ICU discharge (assessed via the Manchester Mobility Score), mean ICU, and post-ICU length of stay (LOS), ventilator days, and in-hospital mortality.

RESULTS: The introduction of the ICU rehabilitation team was associated with a significant increase in mobility at ICU discharge, and this was associated with a significant reduction in ICU LOS (16.9 vs 14.4 days, P = .007), ventilator days (11.7 vs 9.3 days, P < .05), total hospital LOS (35.3 vs 30.1 days, P < .001), and in-hospital mortality (39% vs 28%, P < .05).

CONCLUSION: A quality improvement strategy to promote early and enhanced rehabilitation within this European ICU improved levels of mobility at critical care discharge, and this was associated with reduced ICU and hospital LOS and reduced days of mechanical ventilation.

McWilliams D, Jones, Atkins G, Hodson J, Whitehouse T, Veenith T, Reeves E, Cooper L, Snelson C. Earlier and enhanced rehabilitation of mechanically ventilated patients in critical care: A feasibility randomised controlled trial. J Crit Care. 2018 Apr;44:407-412. doi: 10.1016/j.jcrc.2018.01.001. Epub 2018 Jan 4.

BACKGROUND: Systematic reviews of early rehabilitation within intensive care units have highlighted the need for robust multi-centre randomised controlled trials with longer term follow up. This trial aims to explore the feasibility of earlier and enhanced rehabilitation for patients mechanically ventilated for ≥ 5days and to assess the impact on possible long term outcome measures for use in a definitive trial.

METHODS: Patients admitted to a large UK based intensive care unit and invasively ventilated for ≥ 5days were randomised to the rehabilitation intervention or standard care on a 1:1 basis, stratified by age and SOFA score. The rehabilitation intervention involved a structured programme, with progression along a functionally based mobility protocol according to set safety criteria.

RESULTS: 103 out of 128 eligible patients were recruited into the trial, achieving an initial recruitment rate of 80%. Patients in the intervention arm mobilized significantly earlier (8days vs 10 days, p=0.035), at a more acute phase of illness (SOFA 6 vs 4, p<0.05) and reached a higher level of mobility at the point of critical care discharge (MMS 7 vs 5, p<0.01).

CONCLUSION: We have demonstrated the feasibility of introducing a structured programme of rehabilitation for patients admitted to critical care.

Thrush A[1], Rozek M, Dekerlegand JL. The clinical utility of the functional status score for the intensive care unit (FSS-ICU) at a long-term acute care hospital: a prospective cohort study. Phys Ther. 2012 Dec;92(12):1536-45. doi: 10.2522/ptj.20110412. Epub 2012 Sep 6.

BACKGROUND AND PURPOSE: Long-term acute care hospitals (LTACHs) have emerged for patients requiring medical care beyond a short stay. Minimal data have been reported on functional outcomes in this setting. The purposes of this study were: (1) to measure the clinical utility of the Functional Status Score for the Intensive Care Unit (FSS-ICU) in an LTACH setting and (2) to explore the association between FSS-ICU score and discharge setting.

PARTICIPANTS: Data were obtained from 101 patients (median age=70 years, interquartile range [IQR]=61-78; 39% female, 61% male) who were admitted to an LTACH. Participants were categorized into 1 of 5 groups by discharge setting: (1) home (n=14), (2) inpatient rehabilitation facility (n=26), (3) skilled nursing facility (n=23), (4) long-term care/hospice/expired (n=13), or (5) transferred to a short-stay hospital (n=25).METHODS: Data were prospectively collected from a 38-bed LTACH in the United States over 8 months beginning in September 2010. Functional status was scored using the FSS-ICU within 4 days of admission and every 2 weeks until discharge. The FSS-ICU consists of 5 categories: rolling, supine-to-sit transfers, unsupported sitting, sit-to-stand transfers, and ambulation. Each category was rated from 0 to 7, with a maximum cumulative FSS-ICU score of 35.RESULTS: Cumulative FSS-ICU scores significantly improved from a median (IQR) of 9 (3-17) to 14 (5-24) at discharge. Median (IQR) cumulative discharge FSS-ICU scores were significantly different among the discharge categories: home=28 (22-32), inpatient rehabilitation facility=21 (15-24), skilled nursing facility=14 (8-21), long-term care/hospice/expired=5 (0-11), and transfer to a short-stay hospital=4 (0-7).

DISCUSSION AND CONCLUSIONS: Patients receiving therapy at an LTACH demonstrate significant improvements from admission to discharge using the FSS-ICU. This outcome tool discriminates among discharge settings and successfully documents functional improvements of patients in an LTACH setting.

TEAM Study Investigators. (2015). Early mobilization and recovery in mechanically ventilated patients in the ICU: a bi-national, multi-centre, prospective cohort study. Critical Care, 19(1), 81.

INTRODUCTION: The aim of this study was to investigate current mobilization practice, strength at ICU discharge and functional recovery at 6 months among mechanically ventilated ICU patients.

METHOD: This was a prospective, multi-centre, cohort study conducted in twelve ICUs in Australia and New Zealand. Patients were previously functionally independent and expected to be ventilated for >48 hours. We measured mobilization during invasive ventilation, sedation depth using the Richmond Agitation and Sedation Scale (RASS), co-interventions, duration of mechanical ventilation, ICU-acquired weakness (ICUAW) at ICU discharge, mortality at day 90, and 6-month functional recovery including return to work.

RESULTS: We studied 192 patients (mean age 58.1 ± 15.8 years; mean Acute Physiology and Chronic Health Evaluation (APACHE) (IQR) II score, 18.0 (14 to 24)). Mortality at day 90 was 26.6% (51/192). Over 1,351 study days, we collected information during 1,288 planned early mobilization episodes in patients on mechanical ventilation for the first 14 days or until extubation (whichever occurred first). We recorded the highest level of early mobilization. Despite the presence of dedicated physical therapy staff, no mobilization occurred in 1,079

(84%) of these episodes. Where mobilization occurred, the maximum levels of mobilization were exercises in bed (N = 94, 7%), standing at the bed side (N = 11, 0.9%) or walking (N = 26, 2%). On day three, all patients who were mobilized were mechanically ventilated via an endotracheal tube (N = 10), whereas by day five 50% of the patients mobilized were mechanically ventilated via a tracheostomy tube (N = 18).In 94 of the 156 ICU survivors, strength was assessed at ICU discharge and 48 (52%) had ICU-acquired weakness (Medical Research Council Manual Muscle Test Sum Score (MRC-SS) score <48/60). The MRC-SS score was higher in those patients who mobilized while mechanically ventilated (50.0 ± 11.2 versus 42.0 ± 10.8, P = 0.003). Patients who survived to ICU discharge but who had died by day 90 had a mean MRC score of 28.9 ± 13.2 compared with 44.9 ± 11.4 for day-90 survivors (P <0.0001).

CONCLUSIONS: Early mobilization of patients receiving mechanical ventilation was uncommon. More than 50% of patients discharged from the ICU had developed ICU-acquired weakness, which was associated with death between ICU discharge and day-90.

Trees DW, Smith JM, Hockert S. Innovative mobility strategies for the patient with intensive care unit-acquired weakness: a case report. Phys Ther. 2013 Feb;93(2):237-47. doi: 10.2522/ptj.20110401. Epub 2012 May 10.

BACKGROUND AND PURPOSE: Although the benefits of early mobilization in the intensive care unit (ICU) have been well documented in recent years, the decision-making process and customization of treatment strategies for patients with ICU-acquired weakness have not been well defined in the literature. This case report will describe a patient with ICU-acquired weakness in the long-term acute care hospital (LTACH) setting and mobilization strategies that include novel devices for therapeutic exercise and gait training.

CASE DESCRIPTION: A 73-year-old, active woman underwent a routine cardioversion for atrial fibrillation but developed multiple complications, including sepsis and respiratory failure. The patient spent 3 weeks of limited activity in the ICU and was transferred to our LTACH for continued medical intervention and rehabilitation. A 4-phase graded mobilization program was initiated in the LTACH ICU. Within that program, the physical therapy interventions included partial weight-bearing antigravity strength training with a mobile leg press and gait training with a hydraulic-assist platform walker.

OUTCOME: Before interventions, the patient had severe weakness (Medical Research Council [MRC] sum score of 18/60) and displayed complete dependence for all functioning. She progressed to being able to ambulate 150 ft (1 ft=0.3048 m) using a rolling walker with accompanying strength increases to an MRC sum score of 52/60.

DISCUSSION: This case report describes novel mobility strategies for managing a patient with ICU-acquired weakness. The application of a graded mobilization program using a mobile leg press and a hydraulic-assist platform walker was safe and feasible, and appeared to expedite the patient's recovery process while decreasing the amount of manual lifting for the therapists.

Verceles AC, Wells CL, Sorkin JD, Terrin ML, Beans J, Jenkins T, Goldberg AP. A multimodal rehabilitation program for patients with ICU acquired weakness improves ventilator weaning and discharge home. J Crit Care. 2018 Oct;47:204-210. doi: 10.1016/j.jcrc.2018.07.006.

PURPOSE: To compare the effects of adding a progressive multimodal rehabilitation program to usual care (MRP + UC) versus UC alone on 1) functional mobility, strength, endurance and 2) ventilator weaning and discharge status of patients with ICU-acquired weakness (ICUAW) receiving prolonged mechanical ventilation (PMV).

METHODS: Randomized pilot trial of an individualized MRP + UC versus UC in middle-aged and older ICU survivors with ICUAW receiving PMV. Outcomes compare changes in strength, mobility, weaning success and discharge home from a long-term acute care hospital (LTACH) between the groups.

RESULTS: Eighteen males and 14 females (age 60.3 ± 11.9 years) who received PMV for ≥14 days were enrolled. Despite no significant differences between groups in the changes in handgrip, gait speed, short physical performance battery or 6-min walk distance after treatment, the MRP + UC group had greater weaning success (87% vs. 41%, p < 0.01), and more patients discharged home than UC (53 vs. 12%, p = 0.05).

Post hoc analyses, combining patients based on successful weaning or discharge home, demonstrated significant improvements in strength, ambulation and mobility.

CONCLUSION: The addition of an MRP that improves strength, physical function and mobility to usual physical therapy in LTACH patients with ICUAW is associated with greater weaning success and discharge home than UC alone.

Evidence from 2014:

The importance of monitoring improvement in mobility skills among LTCH patients who require ventilator support at the time of admission is supported by the high prevalence of therapy service provision as part of the treatment plan and the percentage of patients discharged home after an LTCH stay. In a study of 1,419 ventilator-dependent patients from 23 LTCHs with weaning programs (Scheinhorn et al., 2007), physical therapy, occupational therapy, and speech therapy were the three most commonly provided services among 34 procedures, services, and treatments provided during the LTCH admission. The very high frequency of physical (84.8%), occupational (81.5%), and speech (79.7%) therapy reflects use of the rehabilitative model of care adopted by many post-intensive care unit weaning programs, which is important in restoration of function. Improvement in functional status, including mobility and self-care, was noted from admission to discharge. Nearly 30% of all patients discharged alive returned directly home or to assisted living (Scheinhorn et al., 2007). In a sample of 101 patients in an LTCH (three-quarters were ventilator-dependent), median functional status scores using the Functional Status Score for the Intensive Care Unit (FSS-ICU; rolling, supine-to-sit transfers, unsupported sitting, sit-to-stand transfers, and ambulation) improved significantly from admission to discharge, with significant change in all five functional items. Physical therapy interventions focused on early mobilization and consisted of functional tasks, therapeutic exercise, and balance activities that varied according to each patient's individual impairments and limitations. Occupational therapy interventions primarily included cognitive assessment and retraining, activities of daily living-related training, and group therapy sessions for social, behavioral, and physical interventions (Thrush et al., 2012).

Functional improvement is particularly relevant for patients who require ventilator support because these patients traditionally have limited or no mobility because of cardiovascular and pulmonary instability, delirium, sedation, lack of rehabilitation therapy staff, and lack of physician referral (Zanni et al., 2010). A Medicare Payment Advisory Commission (2014) analysis of Medicare Provider Analysis and Review data found that 16 percent of LTCH patients used at least one ventilator-related service in 2012.

Citations

MedPAC "Report to Congress: Medicare Payment Policy" Chapter 11 "Long-term care hospital services." March 2014. Retrieved from http://www.medpac.gov/chapters/Mar14_Ch11.pdf

Scheinhorn, D. J., Hassenpflug, M. S., Votto, J. J., Chao, D. C., Epstein, S. K., Doig, G. S., … Petrak, R. A. (2007). Post-ICU mechanical ventilation at 23 long-term care hospitals: A multicenter outcomes study. *Chest, 131,* 85–93.

Thrush, A., Rozek, M., & Dekerlegand, J. L. (2012). The clinical utility of the Functional Status Score for the Intensive Care Unit (FSS-ICU) at a long-term acute care hospital: A prospective cohort study." *Physical Therapy, 92,* 1536–1545.

Zanni, J. M., Korupolu, R., Fan, E., Pradhan, P., Janjua, K., Palmer, J. B., Needham, D. M. (2010). Rehabilitation therapy and outcomes in acute respiratory failure: An observational pilot project. *Journal of Critical Care, 25,* 254–262.

1a.3. SYSTEMATIC REVIEW(SR) OF THE EVIDENCE (for INTERMEDIATE OUTCOME, PROCESS, OR STRUCTURE PERFORMANCE MEASURES, INCLUDING THOSE THAT ARE INSTRUMENT-BASED) If the evidence is not based on a systematic review go to section 1a.4) If you wish to include more than one systematic review, add additional tables.

2019: Not Applicable. This measure is an outcome measure, so this information is not applicable.

What is the source of the systematic review of the body of evidence that supports the performance measure? A systematic review is a scientific investigation that focuses on a specific question and uses explicit, prespecified scientific methods to identify, select, assess, and summarize the findings of similar but separate studies. It may include a quantitative synthesis (meta-analysis), depending on the available data. (IOM)

☐ Clinical Practice Guideline recommendation (with evidence review)

☐ US Preventive Services Task Force Recommendation

☐ Other systematic review and grading of the body of evidence (*e.g., Cochrane Collaboration, AHRQ Evidence Practice Center*)

☐ Other

2019: Not Applicable. This measure is an outcome measure, so this information is not applicable.

Information in this table is from the 2014 application

Source of Systematic Review: • **Title** • **Author** • **Date** • **Citation, including page number** • **URL**	**I. Centre for Clinical Practice at NICE (UK). (2009). *Rehabilitation after critical illness* (NICE Clinical Guidelines No. 83). Retrieved from** http://www.nice.org.uk/guidance/CG83 **References and citations that support the NICE guidelines can be found at the following location:** http://www.nice.org.uk/guidance/cg83/resources/cg83-critical-illness-rehabilitation-guideline2

Quote the guideline or recommendation verbatim about the process, structure or intermediate outcome being measured. If not a guideline, summarize the conclusions from the SR.	**I. Centre for Clinical Practice at NICE (UK). Rehabilitation after critical illness** (**Full citation:** Centre for Clinical Practice at NICE (UK). (2009). *Rehabilitation after critical illness* (NICE Clinical Guidelines No. 83) Retrieved from http://www.nice.org.uk/guidance/CG83) Key principle of care 1.1 - To ensure continuity of care, healthcare professional(s) with the appropriate competencies1 should coordinate the patient's rehabilitation care pathway. Key elements of the coordination are as follows. • Ensure the short-term and medium-term rehabilitation goals are reviewed, agreed and updated throughout the patient's rehabilitation care pathway. • Ensure the delivery of the structured and supported self-directed rehabilitation manual, when applicable. • Liaise with primary/community care for the functional reassessment at 2–3 months after the patient's discharge from critical care. • Ensure information, including documentation, is communicated between hospitals and to other hospital-based or community rehabilitation services and primary care services. • Give patients the contact details of the healthcare professional(s) on discharge from critical care, and again on discharge from hospital. During the critical care stay 1.2 - During the patient's critical care stay and as early as clinically possible, perform a short clinical assessment to determine the patient's risk of developing physical and non-physical morbidity 1.3 - For patients at risk of physical and non-physical morbidity, perform a comprehensive clinical assessment to identify their current rehabilitation needs. This should include assessments by healthcare professionals experienced in critical care and rehabilitation. 1.5 - The comprehensive clinical assessment and the rehabilitation goals should be collated and documented in the patient's clinical records. 1.6 For patients at risk, start rehabilitation as early as clinically possible, based on the comprehensive clinical assessment and the rehabilitation goals. Before discharge from critical care 1.1.8. - For patients who were previously identified as being at low risk, perform a short clinical assessment before their discharge from critical care to 1.9 - For patients at risk, and patients who started the individualised, structured rehabilitation programme in critical care, perform a comprehensive clinical reassessment to identify their current rehabilitation needs. 1.10 - For patients who were previously identified as being at risk during critical care, the outcomes of the comprehensive reassessment should inform the individualised, structured rehabilitation programme (recommendation 1.1.6). 1.11 - For patients at risk, agree or review and update the rehabilitation goals, based on the comprehensive reassessment. The family and/or carer should also be involved, unless the patient disagrees. Before discharge to home or community care

	1.20 - Before discharging patients who were receiving the individualised structured rehabilitation programme during ward-based care (recommendation 1.1.15): • perform a functional assessment which should include the following physical and non-physical dimensions – physical problems – sensory problems – communication problems – social care or equipment needs – anxiety – depression – post-traumatic stress-related symptoms – behavioural and cognitive problems – psychosocial problems. • assess the impact of the outcomes from the functional assessment on the patient's activities of daily living and participation • based on the functional assessment, review, update and agree the rehabilitation goals with the patient. The family and/or carer should be involved if the patient agrees. 1.21 - If continuing rehabilitation needs are identified from the functional assessment, ensure that before the patient is discharged: • discharge arrangements, including appropriate referrals for the necessary ongoing care, are in place before completing the discharge • all discharge documents are completed and forwarded to the appropriate post-discharge services and the patient • the patient, and/or the family and/or carer as appropriate, is aware of the discharge arrangements and understands them.
Grade assigned to the **evidence** associated with the recommendation with the definition of the grade	
Provide all other grades and definitions from the evidence grading system	
Grade assigned to the **recommendation** with definition of the grade	(**Full citation:** Centre for Clinical Practice at NICE (UK). (2009). *Rehabilitation after critical illness* (NICE Clinical Guidelines No. 83):91. Retrieved from http://www.nice.org.uk/guidance/CG83) Overall, the evidence was of mixed quality. Three out of the seven included studies (Beauchamp et al., 2001; Collen et al., 1991; McKinley & Madronio, 2008) need cautious interpretation because they were graded as low quality based on the Quality Assessment of Diagnostic Accuracy Studies (QUADAS) checklist (with level of evidence '–').

Provide all other grades and definitions from the recommendation grading system	(**Full citation:** Centre for Clinical Practice at NICE (UK). (2009). *Rehabilitation after critical illness* (NICE Clinical Guidelines No. 83):91. Retrieved from http://www.nice.org.uk/guidance/CG83) NICE has produced the guideline based on the best-available evidence, which was presented to a multidisciplinary group of healthcare professionals, patient representatives and carer representatives. The group used its clinical expertise and experience to draft recommendations based on this evidence. It is acknowledged that for rehabilitation after a period of critical care there is a limited evidence base. In some areas there was strong evidence but in other areas the evidence base was weaker or absent. It should be noted that there are many areas of healthcare where there is little or no research-based evidence. Where there is no research-based evidence, standard practice is to use the consensus opinion of the group developing the guideline on what constitutes good practice as the basis for guideline recommendations. For each clinical question the Guideline Development Group (GDG) was presented with a summary of the clinical evidence and, where appropriate, the economic evidence from the studies that were reviewed and appraised. For areas where there was no evidence, the GDG agreed recommendations through informal consensus based on GDG members' experience in the field and their experience in other, related fields such as neurorehabilitation, cardiac rehabilitation and stroke rehabilitation. The committee identified one study (Collen et al. 1991) on the clinical/test utility of an assessment tool for physical morbidity. No studies were identified on screening physical morbidity. The committee identified six studies (Beauchamp et al. 2001; McKinley and Madronio 2008; Stoll et al. 1999; Sukantarat et al. 2007; Twigg et al. 2008; Vedana et al. 2002) on the clinical/test utility of screening tools for non-physical morbidity and one study on assessing cognitive dysfunction (Beauchamp et al. 2001). No studies were identified for screening and/or assessing swallowing and communication problems, and no specific studies were identified on the optimal time to screen for and/or assess physical and non-physical morbidity.
Body of evidence: • Quantity – how many studies? • Quality – what type of studies?	
Estimates of benefit and consistency across studies	
What harms were identified?	
Identify any new studies conducted since the SR. Do the new studies change the conclusions from the SR?	

Source of Systematic Review: • **Title** • **Author** • **Date** • **Citation, including page number** • **URL**	**II. Balas MC, Casey CM, Happ MB. Comprehensive assessment and management of the critically ill. In Boltz M, Capezuti E, Fulmer T, Zwicker D, editor(s). Evidence-based geriatric nursing protocols for best practice. (4th ed., pp. 600-27). New York, NY: Springer.** **II. Comprehensive assessment and management of the critically ill. In: Evidence-based geriatric nursing protocols for best practice.** **(Full citation: Balas MC, Casey CM, Happ MB. Comprehensive assessment and management of the critically ill. In Boltz M, Capezuti E, Fulmer T, Zwicker D, editor(s). Evidence-based geriatric nursing protocols for best practice. (4th ed., pp. 600-27). New York, NY: Springer.)** **AGREE Next Steps Consortium (2009). Appraisal of guidelines for research & evaluation II. Retrieved from http://www.agreetrust.org/?o=1397.** **Adapted from: Melnyck, B. M. & Fineout-Overholt, E. (2005). Evidence-based practice in nursing & health care: A guide to best practice. Philadelphia, PA: Lippincott Williams & Wilkins and Stetler, C.B., Morsi, D., Rucki, S., Broughton, S., Corrigan, B., Fitzgerald, J., et al. (1998). Utilization-focused integrative reviews in a nursing service. Applied Nursing Research, 11(4)**
Quote the guideline or recommendation verbatim about the process, structure or intermediate outcome being measured. If not a guideline, summarize the conclusions from the SR.	**II. Comprehensive assessment and management of the critically ill. In: Evidence-based geriatric nursing protocols for best practice.** **(Full citation:** Balas MC, Casey CM, Happ MB. Comprehensive assessment and management of the critically ill. In Boltz M, Capezuti E, Fulmer T, Zwicker D, editor(s). Evidence-based geriatric nursing protocols for best practice. (4th ed., pp. 600-27). New York, NY: Springer.) **Major Recommendations** Levels of evidence (**I–VI**) are defined at the end of the "Major Recommendations" field. **Interventions and Practices Considered** **Assessment/Evaluation** 1. Comprehensive preadmission assessment: health status, cognitive and functional ability, and social support systems **Major Outcomes Considered** • Functional Status Recommendations Multiple Organ Systems • Encouraging early, frequent mobilization/ambulation Benefits/Harms of Implementing this Guideline Recommendation Potential Benefits • Maintenance/optimization of preadmission functional ability Note: some sections of this clinical guideline have been omitted because they do not apply to this measures setting or focus.
Grade assigned to the **evidence** associated with the recommendation with the definition of the grade	
Provide all other grades and definitions from the evidence grading system	

Grade assigned to the **recommendation** with definition of the grade	**II. Comprehensive assessment and management of the critically ill. In: Evidence-based geriatric nursing protocols for best practice.** **(Full citation:** Balas MC, Casey CM, Happ MB. Comprehensive assessment and management of the critically ill. In Boltz M, Capezuti E, Fulmer T, Zwicker D, editor(s). Evidence-based geriatric nursing protocols for best practice. (4th ed., pp. 600-27). New York, NY: Springer.) See previous section, 1.1 for grades. Grade definitions are: **Level I**: Systematic reviews (integrative/meta-analyses/clinical practice guidelines based on systematic reviews) **Level II**: Single experimental study (randomized controlled trials [RCTs]) **Level III**: Quasi-experimental studies **Level IV**: Non-experimental studies **Level V**: Care report/program evaluation/narrative literature reviews **Level VI**: Opinions of respected authorities/consensus panels
Provide all other grades and definitions from the recommendation grading system	**II. Comprehensive assessment and management of the critically ill. In: Evidence-based geriatric nursing protocols for best practice.** **(Full citation:** Balas MC, Casey CM, Happ MB. Comprehensive assessment and management of the critically ill. In Boltz M, Capezuti E, Fulmer T, Zwicker D, editor(s). Evidence-based geriatric nursing protocols for best practice. (4th ed., pp. 600-27). New York, NY: Springer.) **Level I**: Systematic reviews (integrative/meta-analyses/clinical practice guidelines based on systematic reviews) **Level II**: Single experimental study (randomized controlled trials [RCTs]) **Level III**: Quasi-experimental studies **Level IV**: Non-experimental studies **Level V**: Care report/program evaluation/narrative literature reviews **Level VI**: Opinions of respected authorities/consensus panels
Body of evidence: • Quantity – how many studies? • Quality – what type of studies?	
Estimates of benefit and consistency across studies	
What harms were identified?	
Identify any new studies conducted since the SR. Do the new studies change the conclusions from the SR?	

Source of Systematic Review: • **Title** • **Author** • **Date** • **Citation, including page number** • **URL**	
Quote the guideline or recommendation verbatim about the process, structure or intermediate outcome being measured. If not a guideline, summarize the conclusions from the SR.	
Grade assigned to the **evidence** associated with the recommendation with the definition of the grade	
Provide all other grades and definitions from the evidence grading system	
Grade assigned to the **recommendation** with definition of the grade	
Provide all other grades and definitions from the recommendation grading system	
Body of evidence: • Quantity – how many studies? • Quality – what type of studies?	
Estimates of benefit and consistency across studies	
What harms were identified?	
Identify any new studies conducted since the SR. Do the new studies change the conclusions from the SR?	

1a.4 OTHER SOURCE OF EVIDENCE

If source of evidence is NOT from a clinical practice guideline, USPSTF, or systematic review, please describe the evidence on which you are basing the performance measure.

2019: Not Applicable. This measure is an outcome measure, so this information is not applicable.

From 2014 application:

Functional limitations following critical illness are becoming increasingly prevalent as a result of improving critical care medicine and survival rates (Adler & Malone, 2012). Short- and long-term adverse consequences among critically and chronically ill patients in LTCH and Intensive Care Unit (ICU) settings include severe weakness (Adler & Malone, 2012; Centre for Clinical Practice at NICE [UK], 2009; Dang, 2013; Skinner, Berney, Warrillow, & Denehy, 2009); muscle atrophy (Zanni et al., 2010); connective-tissue shortening (Zanni et al., 2010); loss of bone mass (Dang, 2013); increased risk for blood clots (Dang, 2013); increased risk for pressure ulcers (Dang, 2013); deconditioning (Schweickert & Kress, 2011; Zanni et al., 2010); deficits in self-care and ambulation (Adler & Malone, 2012); functional impairment (Skinner et al., 2009); fatigue (Centre for Clinical Practice at NICE [UK], 2009); and cognitive impairment, including profound and persistent deficits in memory, attention/concentration, and executive function (Brummel et al., 2012; Centre for Clinical Practice at NICE [UK], 2009; Wilcox, Brummel, Archer, Ely, Jackson, & Hopkins, 2013) and the inability to return to work one year after hospital discharge (Dang, 2013; Engel, Needham, Morris, & Gropper, 2013).

To mitigate these adverse consequences, traditional practices of bed rest and immobility have been challenged in recent years, and early mobility and rehabilitation have been increasingly recognized as important to improve patients' long-term functional outcomes (Drolet et al., 2013; Dang, 2013; Skinner at al., 2009; Wilcox et al., 2013; Brummel et al., 2012; Engel et al., 2013; Li et al., 2013), with recovery of function being described as both desirable and possible (Rochester, 2009). The lack of early mobility initiation in intensive care unit settings has also been described as a strong predictor of patient outcomes (Dang, 2013).

Functional improvement is particularly relevant for patients who require ventilator support because these patients traditionally have limited or no mobility because of cardiovascular and pulmonary instability, delirium, sedation, lack rehabilitation therapy staff, and lack of physical referral (Dang, 2013). An increasing body of evidence has reported on the safety and feasibility of early mobilization and rehabilitation of critically ill, but stable, patients in LTCH and intensive care units, with minimal adverse events and risk to the patient (Adler & Malone, 2012; Drolet et al., 2013; Kress, 2009; Schweickert & Kress, 2011; Schweickert et al., 2009; Zanni et al., 2010).

Early mobility and rehabilitation in these settings have been associated with several improved patient outcomes. Reported benefits of early mobility and rehabilitation include: (1) improved strength (Dang, 2013; Li et al., 2013; Schweickert & Kress, 2011) and functional status; (Adler & Malone, 2012; Li et al., 2013; Schweickert & Kress, 2011) (2) earlier achievement of mobilization milestones, such as out of bed mobilization (Adler & Malone, 2012; Morris, 2007); (3) improvement in mobility and self-care function scores from admission to discharge (Li et al., 2013; Scheinhorn et al., 2007); (4) greater incidence of return to functional baseline in mobility and self-care, greater unassisted walking and walking distances, and improved self-reported physical function scores at hospital discharge compared with persons not participating in early mobility and rehabilitation; (Adler & Malone, 2012) (5) enhanced recovery of functional exercise capacity (Dang, 2013); (6) improved self-perceived functional status (Dang, 2013); and (7) reduced physiological and cognitive complications (Dang, 2013) and improved cognitive function (Li et al., 2013). Early mobility and rehabilitation have also been associated with reduced intensive care unit and hospital length of stay (Adler & Malone, 2012; Dang, 2013; Engel et al., 2013; Kress, 2009; Li et al., 2013; Schweickert & Kress, 2011); reduced incidence of delirium and improved patient awareness (Adler & Malone, 2012; Schweickert & Kress, 2011); increased ventilator-free days and improved weaning outcomes (Adler & Malone, 2012; Dang, 2013; Li et al., 2013); greater incidence of discharge home directly after hospitalization compared with patients not receiving early mobilization (Engel et al., 2013; Schweickert et al., 2009); and reduced hospital readmission or death in the year after hospitalization (Adler & Malone, 2012; Li et al., 2013).

Several studies have examined functional improvement among patients in the LTCHs. In a sample of 101 patients in an LTCH (three-quarters were ventilator-dependent), median functional status scores using the Functional Status Score for the Intensive Care Unit (FSS-ICU; rolling, supine-to-sit transfers, unsupported sitting, sit-to-stand transfers, and ambulation) improved significantly from admission to discharge, with significant change in all five functional items. Physical therapy interventions focused on early mobilization and consisted of functional tasks, therapeutic exercise, and balance activities that varied according to each patient's individual impairments and limitations. Occupational therapy interventions primarily included cognitive assessment and retraining, activities of daily living-related training, and group therapy sessions for social, behavioral, and physical interventions. Discharge functional status scores were significantly different across five different discharge destinations (i.e., home, inpatient rehabilitation facility, skilled nursing facility, nursing facility/hospice/expired, and short-stay acute care hospital transfer), highlighting the association of functional status with discharge disposition. A small effect size (0.25) for rehabilitation (physical therapy and occupational therapy) was noted for the entire LTCH sample, and large effect sizes (0.80–0.91) were noted for patients discharged to home, Inpatient Rehabilitation Facilities, or Skilled Nursing Facilities (Thrush et al., 2012).

A separate study of 103 patients with respiratory failure undergoing 1,449 activity events in a respiratory intensive care unit, more than one-half of the activity events were reported to be ambulation, and 40% occurred in intubated, mechanically ventilated patients. At the end of the respiratory intensive care unit stay,

69.4% of survivors ambulated more than 100 feet, 8.2% ambulated less than 100 feet, 15.3% could sit in a chair, 4.7% could sit on the edge of the bed, and 2.4% did not accomplish any of these activities (Bailey et al., 2007).

1a.4.1 Briefly SYNTHESIZE the evidence that supports the measure. A list of references without a summary is not acceptable.

2019: Not Applicable. This measure is an outcome measure, so this information is not applicable.

1a.4.2 What process was used to identify the evidence?

2019: Not Applicable. This measure is an outcome measure, so this information is not applicable.

From 2014 application:

We identified evidence from literature searches using PubMed and the Cumulative Index of Nursing and Allied Health Literature (CINAHL) and in reviews of references cited in the relevant identified studies.

1a.4.3. Provide the citation(s) for the evidence.

2019: Not Applicable. This measure is an outcome measure, so this information is not applicable.

From 2014 application:

1. Adler, J., & Malone, D. (2012). Early mobilization in the intensive care unit: A systematic review. *Cardiopulmonary Physical Therapy Journal, 23*(1), 5–13.

The purpose of this review is to evaluate the literature related to mobilization of the critically ill patient with an emphasis on functional outcomes and patient safety. A search of the scientific literature revealed a limited number of studies that examined the mobilization of critically ill patients in the intensive care unit. However, literature that does exist supports early mobilization and physical therapy as a safe and effective intervention that can have a significant impact on functional outcomes.

2. Bailey, P., Thomsen, G. E., Spuhler, V. J., Blair, R., Jewkes, J., Bezdjian, L., Hopkins, R. O. (2007)Early activity is feasible and safe in respiratory failure patients. Critical Care Medicine, 35(1), 139-145.

The purpose of this study was to investigate if patients in respiratory failure on intensive care units could safely ambulate and complete two bed mobility activates including sit on bed and sit on chair. A total of 1,449 activity events were conducted with 103 patients. The activity events included 233 (16%) sit on bed, 454 (31%) sit in chair, and 762 (53%) ambulate. There were less than <1% of activity-related adverse events.

3. Brummel, N. E., Jackson, J. C., Girard, T. D., Pandharipande, P. P., Schiro, E., Work, B., Ely, E. W. (2012). A combined early cognitive and physical rehabilitation program for people who are critically ill: The Activity and Cognitive Therapy in the Intensive Care Unit (ACT-ICU) Trial. *Physical Therapy, 92,* 1580–1592.

The purpose of this article was to describe a randomized control trial with the purpose to determine the feasibility of early and sustained cognitive rehabilitation paired with physical rehabilitation in patients who are critically ill with respiratory failure or shock. Patients were randomized to groups receiving usual care, physical rehabilitation, or cognitive rehabilitation plus physical rehabilitation. The authors concluded if feasible, these interventions will lay the groundwork for a larger, multicenter trial to determine their efficacy.

4. Dang, S. L. (2013). ABCDEs of ICU: Early mobility. *Critical Care Nursing Quarterly* 36, 163–168.

This issue brief describes the nature and importance of early mobility programs in intensive care units and the impact they can have on positive patient functional outcomes. The author calls for further research to establish and institute policies and protocols on early mobility in the intensive care unit to direct patient care and that will have a positive impact on intensive care unit culture change.

5. Drolet, A., DeJuilio, P., Harkless, S., Henricks, S., Kamin, E., Leddy, E. A., & Williams, S. (2013). Move to improve: The feasibility of using an early mobility protocol to increase ambulation in the intensive and intermediate care settings. *Physical Therapy, 93,* 197–207.

Prolonged bed rest in hospitalized patients leads to deconditioning, impaired mobility, and the potential for longer hospital stays. The initial experience with a nurse-driven mobility protocol suggests that the rate of

patient ambulation in an adult intensive care unit and intensive medical care unit during the first 72 hours of a hospital stay can be increased.

6. Engel, H. J., Needham, D. M., Morris, P. E., & Gropper, M. A. (2013). ICU early mobilization: From recommendation to implementation at three medical centers. *Critical Care Medicine,* 41(Suppl. 9), S69–S80.

The aim of this study was to compare and contrast the process used to implement an early mobility program in intensive care units at three different medical centers and to assess their impact on clinical outcomes in critically ill patients. Instituting a planned, structured intensive care unit early mobility quality improvement project can result in improved outcomes and reduced costs for intensive care unit patients across healthcare systems.

7. Kress, J. P. (2009). Clinical trials of early mobilization of critically ill patients. [Review]. Critical Care Medicine, 37(Suppl 10).

   This briefing paper describes a review of the literature focused on mechanically ventilated intensive care unit patients and the positive outcomes that are associated with early mobilization. The author describes how a multidisciplinary team approach is needed for patient care. The author calls for further prospective studies of early mobilization to evaluate further the mobilizing of mechanically ventilated intensive care unit patients.

8. Li, Z., Peng, X., Zhu, B., Zhang, Y., & Xi, X. (2013). Active mobilization for mechanically ventilated patients: A systematic review. *Archives of Physical Medicine and Rehabilitation, 94,* 551–561.

This study investigates the effectiveness and safety of active mobilization on improving physical function. Active mobilization appears to have a positive effect on physical function and hospital outcomes in mechanical ventilation patients. Early active mobilization protocols may be initiated safely in the intensive care unit and continued in post-intensive care unit. However, the current available studies have great heterogeneity and limited methodologic quality. Further research is needed to provide more robust evidence to support the effectiveness and safety of active mobilization.

9. Morris, P. E. (2007). Moving our critically ill patients: mobility barriers and benefits. Critical Care Clinics, 23(1), 1-20. doi: 10.1016/j.ccc.2006.11.003

Diagnosis and resuscitation for critically ill patients have improved in the last 25 years, and survival has also increased. With improvements in mortality, the field of critical care has seen increased opportunities to improve posthospital quality of life for survivors of critical illness. This article focuses particularly on how mobilization may improve quality of life for patients.

10. Rochester, C. L. (2009). Rehabilitation in the intensive care unit. *Seminars in Respiratory and Critical Care Medicine, 30,* 656–669.

Critical illness has many devastating sequelae, including profound neuromuscular weakness and psychological and cognitive disturbances that frequently result in long-term functional impairments. Studies conducted to date suggest that such intensive care unit-based rehabilitation is feasible, safe, and effective for carefully selected patients. Further research is needed to identify the optimal patient candidates and procedures and for providing rehabilitation in the intensive care unit.

11. Scheinhorn, D. J., Hassenpflug, M. S., Votto, J. J., Chao, D. C., Epstein, S. K., Doig, G. S., … Petrak, R. A. (2007). Post-ICU mechanical ventilation at 23 long-term care hospitals: A multicenter outcomes study. *Chest, 131,* 85–93.

This multicenter study was undertaken to characterize the population of ventilator-dependent patients admitted to long-term care hospitals (LTCHs) with weaning programs, and to report treatments, complications, weaning outcome, discharge disposition, and survival in these patients. Patients admitted to LTCHs for weaning attempts were elderly, with acute-on-chronic diseases, and continued to require considerable medical interventions and treatments. The frequency and type of complications were not surprising following prolonged and aggressive intensive care unit interventions. In the continuum of critical care medicine, more

than half of ventilator-dependent survivors of catastrophic illness transferred from the intensive care unit were successfully weaned from prolonged mechanical ventilation in the setting of an LTCH.

12. Schweickert, W. D., Pohlman, M. C., Pohlman, A. S., Nigos, C., Pawlik, A. J., Esbrook, C. L., . . . Kress, J. P. (2009). Early physical and occupational therapy in mechanically ventilated, critically ill patients: a randomised controlled trial. Lancet, 373(9678), 1874-1882. doi: 10.1016/s0140-6736(09)60658-9

This randomized control trial assessed the efficacy of combining daily interruption of sedation with physical and occupational therapy on functional outcomes in 104 patients receiving mechanical ventilation in intensive care units in two university hospitals. Patients were randomly assigned by computer-generated, permuted block randomization to early exercise and mobilization (physical and occupational therapy) during periods of daily interruption of sedation (intervention; n=49) or to daily interruption of sedation with therapy as ordered by the primary care team (control; n=55). Return to independent functional status at hospital discharge occurred in 29 (59%) patients in the intervention group compared with 19 (35%) patients in the control group (p=0.02; odds ratio 2.7). Patients in the intervention group had shorter duration of delirium (median 2.0 days vs. 4.0 days p=0.02), and more ventilator-free days (23.5 days vs. 21.1 days; p=0.05) during the 28-day follow-up period than did controls. There was one serious adverse event in 498 therapy sessions (desaturation less than 80%). Discontinuation of therapy as a result of patient instability occurred in 19 (4%) of all sessions, most commonly for perceived patient-ventilator asynchrony. A strategy for whole-body rehabilitation-consisting of interruption of sedation and physical and occupational therapy in the earliest days of critical illness-was safe and well tolerated, and resulted in better functional outcomes at hospital discharge, a shorter duration of delirium, and more ventilator-free days compared with standard care.

13. Schweickert, W. D., & Kress, J. P. (2011). Implementing early mobilization interventions in mechanically ventilated patients in the ICU. *Chest, 140,* 1612–1617.

As intensive care unit survival continues to improve, clinicians are faced with short- and long-term consequences of critical illness. Deconditioning and weakness have become common problems in survivors of critical illness requiring mechanical ventilation. Recent literature, mostly from a medical population of patients in the intensive care unit, has challenged the patient care model of prolonged bed rest. Instead, the feasibility, safety, and benefits of early mobilization of mechanically ventilated intensive care unit patients have been reported in recent publications. The benefits of early mobilization include reductions in length of stay in the intensive care unit and hospital as well as improvements in strength and functional status. Such benefits can be accomplished with a remarkably acceptable patient safety profile. The importance of interactions between mind and body are highlighted by these studies, with improvements in patient awareness and reductions in intensive care unit delirium being noted.

14. Skinner, E. H., Berney, S., Warrillow, S., & Denehy, L. (2009). Development of a physical function outcome measure (Pfit) and a pilot exercise training protocol for use in intensive care. *Critical Care and Resuscitation, 11,* 110–115.

The study aim is to develop an outcome measure as a basis for prescribing and evaluating rehabilitation in the critically ill, and to measure its reliability and responsiveness to change. The study also aimed to assess the feasibility and safety of a pilot exercise training protocol in an intensive care unit. The PFIT is a reliable and responsive outcome measure, and the pilot training protocol was safe and feasible. As exercise may attenuate weakness and functional impairment, the PFIT can be used to prescribe and evaluate exercise and mobilisation.

15. Thrush, A., Rozek, M., & Dekerlegand, J. L. (2012). The clinical utility of the Functional Status Score for the Intensive Care Unit (FSS-ICU) at a long-term acute care hospital: A prospective cohort study." *Physical Therapy, 92,* 1536–1545.

Long-term acute hospitals (LTCHs) have emerged for patients requiring medical care beyond a short stay. Minimal data have been reported on functional outcomes in this setting. The purposes of this study were: (1) to measure the clinical utility of the Functional Status Score for the Intensive Care Unit (FSS-ICU) in an LTCH setting and (2) to explore the association between FSS-ICU score and discharge setting. Patients receiving therapy at an LTCH demonstrate significant improvements from admission to discharge using the FSS-ICU. This

outcome tool discriminates among discharge settings and successfully documents functional improvements of patients in an LTCH.

16. Wilcox, M. E., Brummel, N. E., Archer, K., Ely, E. W., Jackson, J. C., & Hopkins, R. O. (2013). Cognitive dysfunction in ICU patients: risk factors, predictors, and rehabilitation interventions. [Review]. Critical Care Medicine, 41(9 Suppl. 1).

In contrast to other clinical outcomes, long-term cognitive function in critical care survivors has not been deeply studied. In this narrative review, we summarize the existing literature on the prevalence, mechanisms, risk factors, and prediction of cognitive impairment after surviving critical illness. Depending on the exact clinical subgroup, up to 100% of critical care survivors may suffer some degree of long-term cognitive impairment at hospital discharge; in approximately 50%, decrements in cognitive function will persist years later. Although the mechanisms of acquiring this impairment are poorly understood, several risk factors have been identified. Unfortunately, no easy means of predicting long-term cognitive impairment exists. Despite this barrier, research is ongoing to test possible treatments for cognitive impairment. The potential role of exercise on cognitive recovery is an exciting area of exploration.

17. Zanni, J. M., Korupolu, R., Fan, E., Pradhan, P., Janjua, K., Palmer, J. B., … Needham, D. M. (2010). Rehabilitation therapy and outcomes in acute respiratory failure: An observational pilot project. *Journal of Critical Care, 25,* 254–262.

The aim of this study was to describe the frequency, physiologic effects, safety, and patient outcomes associated with traditional rehabilitation therapy in patients who require mechanical ventilation. This pilot project illustrated important barriers to providing rehabilitation to mechanically ventilated patients in an intensive care unit and impairments in strength, range of motion, and functional outcomes at hospital discharge.

1b. Performance Gap

Demonstration of quality problems and opportunity for improvement, i.e., data demonstrating:

- considerable variation, or overall less-than-optimal performance, in the quality of care across providers; and/or
- Disparities in care across population groups.

1b.1. Briefly explain the rationale for this measure *(e.g., how the measure will improve the quality of care, the benefits or improvements in quality envisioned by use of this measure)*

If a COMPOSITE (e.g., combination of component measure scores, all-or-none, any-or-none), SKIP this question and answer the composite questions.

Patients in LTCHs present with clinically complex conditions. In addition to having complex medical care needs for an extended period of time, LTCH patients often have functional limitations due to the nature of their conditions, as well as deconditioning due to prolonged bed rest and treatment requirements (for example, ventilator use). These patients are therefore at high risk for functional decline during the LTCH stay that is both condition-related and iatrogenic (i.e., related to medical treatment).

According to the Centers for Disease Control and Prevention (CDC), more than 300,000 patients receive mechanical ventilation in the United States each year (CDC 2014). These patients are at increased risk for infections, such as pneumonia and sepsis, as well as other serious complications including pulmonary edema, pulmonary embolism, and death (Esteban et al 2002; Klompas et. al 2011). These complications can lead to longer stays in the intensive care unit and hospital, increased health care costs and increased risk of disability (or death) (CDC 2014). The estimated mortality rate in patients aged 85 years and older with acute lung injury on mechanical ventilation is 60 percent (Rubenfeld et al. 2005).

A Medicare Payment Advisory Commission analysis of Medicare data found that 16 percent of LTCH patients used at least one ventilator-related service in 2012 (MedPAC 2014). In fiscal year 2012, MS-LTC-DRG 207, a diagnosis-related group that refers to respiratory diagnosis with ventilator support for 96 or more hours,

represented the most frequently occurring diagnosis among LTCH patients, at 11.3 percent of all LTCH discharges (MedPAC 2014) and MS-LTC-DRG-4, a diagnosis-related group that refers to tracheostomy with ventilator support for 96 or more hours or primary diagnosis except face, mouth, and neck without major OR procedure, represented an additional 1.3 percent of all LTCH discharges. Together, the two diagnosis-related groups account for a total of nearly 18,000 discharges. Furthermore, the number of ventilated patients in LTCHs has increased. The number of patients discharged with a respiratory diagnosis with ventilator support for 96 or more hours increased 7.4 percent between 2008 and 2011 (MedPAC 2014).

Functional improvement is particularly relevant for patients who require ventilator support because these patients have traditionally had limited mobility due to cardiovascular and pulmonary instability, delirium, sedation, lack of rehabilitation therapy staff, and lack of physician referral (Zanni et. al 2010).

In describing the importance of functional status, the National Committee on Vital and Health Statistics Subcommittee on Health (2001) noted, “Information on functional status is becoming increasing essential for fostering healthy people and a health population. Achieving optimal health and well-being for Americans requires an understanding across the life space of the effects of people’s health conditions on their ability to do basic activities and participate in life situations, in other words, their functional status."

This quality measure will inform LTCH providers about opportunities to improve care in the area of function and strengthen incentives for quality improvement related to patient function.

Citations

Centers for Disease Control and Prevention (CDC). Ventilator-Associated Event (VAE). January 2014. http://www.cdc.gov/nhsn/PDFs/pscManual/10-VAE_FINAL.pdf.

Esteban, A., A. Anzueto, et al. (2002). Characteristics and outcomes in adult patients receiving mechanical ventilation: a 28-day international study. JAMA 287(3): 345-355.

Klompas, M., Y. Khan, et al. (2011). Multicenter Evaluation of a Novel Surveillance Paradigm for Complications of Mechanical Ventilation. PLoS ONE 6(3): e18062.

MedPAC. Report to Congress: Medicare Payment Policy. Chapter 11: Long-term care hospital services. March 2014. http://www.medpac.gov/chapters/Mar14_Ch11.pdf.

Rubenfeld, G. D., E. Caldwell, et al. (2005). Incidence and outcomes of acute lung injury. N Engl J Med 353(16): 1685-1693.

Zanni, J. M., Korupolu, R., Fan, E., Pradhan, P., Janjua, K., Palmer, J. B., Needham, D. M. (2010). Rehabilitation therapy and outcomes in acute respiratory failure: an observational pilot project. Journal of Critical Care, 25(2), 254-262. doi: 10.1016/j.jcrc.2009.10.010

National Committee on Vital and Health Statistics Subcommittee on Health. (2001). Classifying and Reporting Functional Status.

1b.2. Provide performance scores on the measure as specified (current and over time) at the specified level of analysis. *(This is required for maintenance of endorsement. Include mean, std dev, min, max, interquartile range, scores by decile. Describe the data source including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities include.) This information also will be used to address the sub-criterion on improvement (4b1) under Usability and Use.*

We provide trend data starting with Quarter 3 of 2016 until Quarter 2 of 2018. Performance scores were calculated using rolling four quarters of data as well as scores by quarter that were conducted using the LTCH CARE Data Set data. Note that providers with < 20 stays during the four rolling quarter periods are excluded.

Quality measure score distributions over the following four rolling quarter periods (n=# of providers):

1. Quarter 3 2016 – Quarter 2 2017 (July 1, 2016 – June 30, 2017) (n=300)

2. Quarter 4 2016 – Quarter 3 2017 (October 1, 2016 – September 30, 2017) (n=297)

3. Quarter 1 2017 – Quarter 4 2017 (January 1, 2017 – December 31, 2017) (n=294)

4. Quarter 2 2017 – Quarter 1 2018 (April 1, 2017 – March 31, 2018) (n=292)

5. Quarter 3 2017 – Quarter 2 2018 (July 1, 2017 – June 30, 2018) (n=284)

Quality measure score distributions by quarter between July 1, 2016 – June 30, 2018 (8 quarters):

1. Quarter 3 2016 (n=386)

2. Quarter 4 2016 (n=383)

3. Quarter 1 2017 (n=383)

4. Quarter 2 2017 (n=378)

5. Quarter 3 2017 (n=374)

6. Quarter 4 2017 (n=362)

7. Quarter 1 2018 (n=358)

8. Quarter 2 2018 (n=355)

Quality measure score distributions over time were stable across the 5 rolling four quarters. The mean national scores ranged from 9.0 to 9.2. Quality measure scores by decile show variations in quality measure scores across LTCHs. The interquartile range for the rolling four quarters data ranged from 4.1 to 4.5 mobility units. Across eight quarters (Q3 2016 – Q2 2018), quality measure score distributions showed variation in LTCH outcomes. Between Q3 2016 to Q2 2018, the overall mean decreased marginally from 9.2 to 9.1.

Rolling Four Quarters:

1) Quarter 3 2016 – Quarter 2 2017

Facilities: 300

Mean score: 9.2

Standard deviation: 3.1

Interquartile range: 4.1

Minimum: -0.2

1st decile: 5.3

2nd decile: 6.5

3rd decile: 7.5

4th decile: 8.3

5th decile: 9.1

6th decile: 9.8

7th decile: 10.7

8th decile: 11.5

9th decile: 13.3

Maximum: 20.3

2) Quarter 4 2016 – Quarter 3 2017

Facilities: 297

Mean score: 9.0

Standard deviation: 3.2

Interquartile range: 4.1

Minimum: 1.0

1st decile: 5.1

2nd decile: 6.5

3rd decile: 7.3

4th decile: 8.0

5th decile: 8.6

6th decile: 9.5

7th decile: 10.4

8th decile: 11.5

9th decile: 13.3

Maximum: 20.3

3) Quarter 1 2016 – Quarter 4 2017

Facilities: 294

Mean score: 9.0

Standard deviation: 3.3

Interquartile range: 4.1

Minimum: 2.1

1st decile: 5.2

2nd decile: 6.5

3rd decile: 7.1

4th decile: 7.9

5th decile: 8.7

6th decile: 9.5

7th decile: 10.4

8th decile: 11.6

9th decile: 13.3

Maximum: 22.9

4) Quarter 2 2017 – Quarter 1 2018

Facilities: 292

Mean score: 9.1

Standard deviation: 3.4

Interquartile range: 4.5

Minimum: 1.4

1st decile: 5.2

2nd decile: 6.4

3rd decile: 7.1

4th decile: 7.8

5th decile: 8.6

6th decile: 9.9

7th decile: 10.8

8th decile: 12.0

9th decile: 13.4

Maximum: 22.3

5) Quarter 3 2017 – Quarter 2 2018

Facilities: 284

Mean score: 9.0

Standard deviation: 3.2

Interquartile range: 4.3

Minimum: 1.5

1st decile: 5.0

2nd decile: 6.5

3rd decile: 7.3

4th decile: 7.7

5th decile: 8.7

6th decile: 9.7

7th decile: 10.4

8th decile: 11.9

9th decile: 13.0

Maximum: 23.6

Quality Measure Score Distributions by Quarter

1) Quarter 3 2016

Facilities: 386

Mean score: 9.4

Standard deviation: 6.1

Interquartile range: 6.5

Minimum: -6.1

Maximum: 36.0

2) Quarter 4 2016

Facilities: 383

Mean score: 9.1

Standard deviation: 6.5

Interquartile range: 5.9

Minimum: -6.2

Maximum: 38.2

3) Quarter 1 2017

Facilities: 383

Mean score: 9.5

Standard deviation: 6.0

Interquartile range: 6.3

Minimum: -3.3

Maximum: 36.5

4) Quarter 2 2017

Facilities: 378

Mean score: 9.8

Standard deviation: 6.0

Interquartile range: 6.4

Minimum: -0.6

Maximum: 34.1

5) Quarter 3 2017

Facilities: 374

Mean score: 9.5

Standard deviation: 5.8

Interquartile range: 6.5

Minimum: -3.5

Maximum: 32.2

6) Quarter 4 2017

Facilities: 362

Mean score: 9.0

Standard deviation: 6.1

Interquartile range: 5.9

Minimum: -2.8

Maximum: 37.0

7) Quarter 1 2018

Facilities: 358

Mean score: 9.7

Standard deviation: 5.6

Interquartile range: 6.0

Minimum: -17.0

Maximum: 38.0

8) Quarter 2 2018

Facilities: 355

Mean score: 9.3

Standard deviation: 5.8

Interquartile range: 6.0

Minimum: -3.5

Maximum: 32.6

Note: Scores are reported as units of change in mobility; Providers with < 20 stays during the four rolling quarter periods are excluded.

Source: RTI analysis of LTCH CARE Data Set July 2016 – June 2018 (Program reference: 2632_03).

1b.3. If no or limited performance data on the measure as specified is reported in 1b2, then provide a summary of data from the literature that indicates opportunity for improvement or overall less than optimal performance on the specific focus of measurement.

An increasing body of evidence has reported on the safety and feasibility of early mobilization and rehabilitation of critically ill but stable patients in LTCH and intensive care units with minimal adverse events and risk to the patient (Adler & Malone, 2012; Drolet et al., 2013; Kress, 2009 ; Schweickert & Kress, 2011; Schweickert et al., 2009 ; Zanni et al., 2010). Early mobility and rehabilitation in these settings have been associated with several improved patient outcomes, such as (1) improved strength (Dang, 2013; Li, Peng, Zhu, Zhang, & Xi, 2013; Schweickert & Kress, 2011) and functional status (Adler & Malone, 2012; Li et al., 2013; Schweickert & Kress, 2011); (2) earlier achievement of mobilization milestones, such as out-of-bed mobilization (Adler & Malone, 2012; Morris, 2007); (3) improvement in mobility and self-care function scores from admission to discharge (Li et al., 2013; Scheinhorn et al., 2007); (4) greater incidence of return to functional baseline in mobility and self-care, greater unassisted walking and walking distances, and improved self-reported physical function scores at hospital discharge compared with persons not participating in early mobility and rehabilitation (Adler & Malone, 2012); (5) enhanced recovery of functional exercise capacity (Dang, 2013); (6) improved self-perceived functional status (Dang, 2013); (7) reduced physiological and cognitive complications (Dang, 2013); and (8) improved cognitive function (Li et al., 2013). Early mobility and rehabilitation have also been associated with (1) reduced ICU and hospital length of stay (Adler & Malone, 2012; Dang, 2013; Engel, Needham, Morris, & Gropper, 2013; Kress, 2009; Li et al., 2013; Schweickert & Kress, 2011), (2) reduced incidence of delirium and improved patient awareness (Adler & Malone, 2012; Schweickert & Kress, 2011), (3) increased ventilator-free days and improved weaning outcomes (Adler & Malone, 2012; Dang, 2013; Li et al., 2013), (4) greater incidence of discharge home directly after hospitalization compared with patients not receiving early mobilization (Engel et al., 2013; Schweickert et al., 2009), and (5) reduced hospital readmission or death in the year after hospitalization (Adler & Malone, 2012; Li et al., 2013).

Mobility activities that are feasible to assess in LTCH and intensive care units include bed mobility, sitting at the edge of the bed, transferring from bed to chair, sitting in a chair, out-of-bed mobility, standing, and ambulation (Adler & Malone, 2012; Bailey et al., 2007; Morris, 2007; Schweickert et al., 2009). In a sample of 103 patients with respiratory failure undergoing 1,449 activity events in a respiratory intensive care unit, more than one-half of the activity events were reported to be ambulation, and 40% of the activity events occurred in intubated, mechanically ventilated patients at the end of the respiratory intensive care unit stay. Moreover, 69.4% of survivors ambulated more than 100 feet, 8.2% ambulated less than 100 feet, 15.3% could sit in a chair, 4.7% could sit on the edge of the bed, and 2.4% did not accomplish any of these activities (Bailey et al., 2007).

As noted in our summary of evidence that links rehabilitation therapy services and functional outcomes, LTCHs and intensive care units are testing new approaches to mobilize patients who require ventilator support. However, many patients are not mobilized early. The American Thoracic Society/American College of Chest Physicians' Clinical Practice Guideline focused on liberation from mechanical ventilation in critically ill adults recommends early mobilization for these patients (Girard et al., 2017).

Citations

Adler, J., & Malone, D. (2012). Early mobilization in the intensive care unit: A systematic review. Cardiopulmonary Physical Therapy Journal, 23(1), 5–13.

Bailey, P., Thomsen, G. E., Spuhler, V. J., Blair, R., Jewkes, J., Bezdjian, L., … Hopkins, R. O. (2007). Early activity is feasible and safe in respiratory failure patients. Critical Care Medicine, 35, 139–145.

Dang, S. L. (2013). ABCDEs of ICU: Early mobility. Critical Care Nursing Quarterly, 36, 163–168.

Drolet, A., DeJuilio, P., Harkless, S., Henricks, S., Kamin, E., Leddy, E. A., & Williams, S. (2013). Move to improve: The feasibility of using an early mobility protocol to increase ambulation in the intensive and intermediate care settings. Physical Therapy, 93, 197–207.

Engel, H. J., Needham, D. M., Morris, P. E., & Gropper, M. A. (2013). ICU early mobilization: From recommendation to implementation at three medical centers. Critical Care Medicine, 41(Suppl. 9), S69–S80. doi:10.1097/CCM.0b013e3182a240d5

Girard TD, Alhazzani W, Kress JP, Ouellette DR, Schmidt GA, Truwit JD, Burns SM, Epstein SK, Esteban A, Fan E, Ferrer M, Fraser GL, Gong MN, Hough CL, Mehta S, Nanchal R, Patel S, Pawlik AJ, Schweickert WD, Sessler CN, Strøm T, Wilson KC, Morris PE; ATS/CHEST Ad Hoc Committee on Liberation from Mechanical Ventilation in Adults. An Official American Thoracic Society/American College of Chest Physicians Clinical Practice Guideline: Liberation from Mechanical Ventilation in Critically Ill Adults. Rehabilitation Protocols, Ventilator Liberation Protocols, and Cuff Leak Tests. Am J Respir Crit Care Med. 2017 Jan 1;195(1):120-133. doi: 10.1164/rccm.201610-2075ST.

Kress, J. P. (2009). Clinical trials of early mobilization of critically ill patients. Critical Care Medicine, 37, S442–S447.

Li, Z., Peng, X., Zhu, B., Zhang, Y., & Xi, X. (2013). Active mobilization for mechanically ventilated patients: A systematic review. Archives of Physical Medicine and Rehabilitation, 94, 551–561.

Morris, P. E. (2007). Moving our critically ill patients: Mobility barriers and benefits. Critical Care Clinics, 23, 1–20.

Scheinhorn, D. J., Hassenpflug, M. S., Votto, J. J., Chao, D. C., Epstein, S. K., Doig, G. S., ... Petrak, R. A. (2007). Post-ICU mechanical ventilation at 23 long-term care hospitals: A multicenter outcomes study. Chest, 131, 85–93. doi:10.1378/chest.06-1081

Schweickert, W. D., Pohlman, M. C., Pohlman, A. S., Nigos, C., Pawlik, A. J., Esbrook, C. L., ... Kress, J. P. (2009). Early physical and occupational therapy in mechanically ventilated, critically ill patients: A randomised controlled trial. Lancet, 373(9678), 1874–1882. doi:10.1016/s0140-6736(09)60658-9

Schweickert, W. D., & Kress, J. P. (2011). Implementing early mobilization interventions in mechanically ventilated patients in the ICU. Chest, 140, 1612–1617. doi:10.1378/chest.10-2829

Zanni, J. M., Korupolu, R., Fan, E., Pradhan, P., Janjua, K., Palmer, J. B., Needham, D. M. (2010). Rehabilitation therapy and outcomes in acute respiratory failure: An observational pilot project. Journal of Critical Care, 25, 254–262. doi:10.1016/j.jcrc.2009.10.010

1b.4. Provide disparities data from the measure as specified (current and over time) by population group, e.g., by race/ethnicity, gender, age, insurance status, socioeconomic status, and/or disability. *(This is required for maintenance of endorsement. Describe the data source including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities included.) For measures that show high levels of performance, i.e., "topped out", disparities data may demonstrate an opportunity for improvement/gap in care for certain sub-populations. This information also will be used to address the sub-criterion on improvement (4b1) under Usability and Use.*

We used the LTCH CARE Data Set to examine whether there may be disparities in care for population groups related to this measure. Disparities for certain population groups would indicate gaps in care and opportunities for improvement. The LTCH CARE Data Set included 411 LTCHs who discharged 39,338 patients requiring ventilator support from July 1, 2016 to June 30, 2018.

We address the issue of disparities for this measure by examining whether there are differences in functional outcomes for population groups that may reflect experience disparities in care, such as for population groups with social risk factors.

We examined whether 4 social risk factors affected performance measure scores. The social risk factors are: 1) payer source (patient-level variable); 2) marriage status (patient-level variable); 3) race (patient-level variable); and 4) ethnicity (patient-level variable). Details about how we obtained and calculated this disparities data is available in Sections 1.2 and 1.8 of the Testing form.

We conducted the following analyses to examine the effect of the 4 social risk factors:

1) We calculated the percentage of stays for each social risk factor population group;
2) We calculated the observed change in mobility score for each social risk factor population group;
3) We added indicators for each social risk factor group to our risk adjustment model and estimated the coefficients for each group (relative to the reference group) in the model;
4) We examined the indicators for each social risk factor over time by quarter in our risk adjustment model to examine whether there may be trends for population groups.

Below is a summary of these analyses and results. For more information on disparities in change in mobility related to race/ethnicity, marriage status, and payer source, please refer to the risk adjustment analyses in the Testing form. Tables and graphics are able to be inserted into the NQF Testing form, unlike this Measure Information form, so we direct readers to Section 2b3.4b of the Testing form for the results presented below in a more readable format (Tables 13, 14, and 15 specifically).

1) The Distribution of Social Risk Factor Patient Population Groups:

We found that 31.2% of patients were Medicaid beneficiaries, 66.0% of patients were white, and 39.7% were not currently married.

2) Observed Change in Mobility Score by Social Risk Factor:

The unadjusted mean change in mobility score varied by race/ethnicity and payer source. Black (6.6) and Asian (6.3) patients had lower unadjusted change in mobility scores than White (9.5) patients. Medicaid (8.0) patients, who were likely dual-eligible, had lower unadjusted change scores than both patients with Medicare (8.4) and private insurance (9.5). Patients not currently married (8.4) also had lower unadjusted changes scores than patients who are married (9.1). Mean change in mobility scores were similar when examining ethnicity (8.7 for non-Hispanic and 8.5 for Hispanic).

3) Estimated Effect (Coefficient Values) for Each Social Risk Factor (24 Months: July 1, 2016 through June 30, 2018)

Each social risk factor was then added to our Generalized Linear regression model to get estimated regression coefficients which represent the effect of each individual factor on change in mobility relative to the reference group. The dependent variable was the change in mobility score for each patient, calculated as the difference between the discharge mobility score and admission mobility score. For example, a coefficient value (ß) of -0.5 for Black patients would be interpreted to mean that, on average, these patients had a change in mobility score that was 0.5 mobility units less than White patients (the reference group).

Compared to patients who were White, Black (ß = -1.6300; $p < 0.0001$) and Asian (ß = -0.4419; $p = 0.2763$) patients had lower change in mobility scores though this was not significant for Asian patients. American Indian/Alaskan Native (ß = 2.6362; $p = < 0.0001$) and Hispanic (ß = 1.3576; $p = < 0.0001$) patients had on average higher mobility changes than White patients, while patients not currently married (ß = -0.1537; $p = 0.1540$) had slightly lower changes, on average compared to patients who were not currently married. Compared to patients with private insurance, patients on Medicaid (ß = -0.7405; $p = 0.0021$ for managed care and ß = -0.3996; $p = 0.0497$ for fee-for-service) had on average lower change in mobility scores.

4) Estimated Coefficient Values for Each Social Risk Factor (by rolling four quarters)

The analysis described above examining each social risk factor's effect on change in mobility for patients requiring ventilator support was then performed by rolling four quarters within the 24-month period to examine possible trends over time. The patients included in each rolling four quarter period and detailed results are provided below.

We observed some coefficient differences across this 24-month period for one social risk factor. Specifically, patients on Medicaid managed care went from having significantly lower mobility scores (ß = -1.1333; $p = 0.0016$) than patients on private insurance to not being significantly different (ß = -0.2979; $p = 0.3854$) towards the end of the 24-month period. On average, Black patients (coeff. range = -1.3418 to -1.9278) consistently had

significantly lower mobility change scores than White patients. Hispanic patients and patients not married had consistent change in mobility scores throughout the 24-month period.

Our testing of social risk factors and their relationships to patients' change in mobility scores indicate that some factors (Medicaid payer, Black race) were tied to lower mobility change scores while others (Hispanic ethnicity, American Indian or Alaska Native race) were tied to higher mobility change scores. We believe that continued monitoring of potential disparities in functional outcomes is critical.

Breakdown of patients discharged within each rolling four quarter period:

Q3 2016 – Q2 2017 (July 1, 2016 – June 30, 2017) = 18,835

Q4 2016 – Q3 2017 (October 1, 2016 – September 30, 2017) = 18,772

Q1 2017 – Q4 2017 (January 1, 2017 – December 31, 2017) = 18,747

Q2 2017 – Q1 2018 (April 1, 2017 – March 31, 2018) = 18,853

Q3 2017 – Q2 2018 (July 1, 2017 – June 30, 2018) = 18,709

Race/Ethnicity (reference = White)

Black

- Q3 2016 – Q2 2017: estimate = -1.3418; SE = 0.20; p-value < 0.0001
- Q4 2016 – Q3 2017: estimate = -1.5346; SE = 0.20; p-value < 0.0001
- Q1 2017 – Q4 2017: estimate = -1.4825; SE = 0.19; p-value < 0.0001
- Q2 2017 – Q1 2018: estimate = -1.5629; SE = 0.20; p-value < 0.0001
- Q3 2017 – Q2 2018: estimate = -1.9278; SE = 0.19; p-value < 0.0001

Asian

- Q3 2016 – Q2 2017: estimate = -0.4226; SE = 0.62; p-value = 0.4968
- Q4 2016 – Q3 2017: estimate = -0.3829; SE = 0.62; p-value = 0.5396
- Q1 2017 – Q4 2017: estimate = -0.1250; SE = 0.61; p-value = 0.8374
- Q2 2017 – Q1 2018: estimate = -0.6720; SE = 0.60; p-value = 0.2620
- Q3 2017 – Q2 2018: estimate = -0.6864; SE = 0.57; p-value = 0.2297

American Indian or Alaska Native

- Q3 2016 – Q2 2017: estimate = 2.1083; SE = 0.91; p-value = 0.0203
- Q4 2016 – Q3 2017: estimate = 1.6474; SE = 0.96; p-value = 0.0858
- Q1 2017 – Q4 2017: estimate = 1.7886; SE = 0.93; p-value = 0.0551
- Q2 2017 – Q1 2018: estimate = 2.2172; SE = 0.90; p-value = 0.0134
- Q3 2017 – Q2 2018: estimate = 3.1504; SE = 0.90; p-value = 0.0004

Native Hawaiian or Pacific Islander

- Q3 2016 – Q2 2017: estimate = 1.4959; SE = 1.59; p-value = 0.3477
- Q4 2016 – Q3 2017: estimate = 1.6814; SE = 1.62; p-value = 0.2986
- Q1 2017 – Q4 2017: estimate = 2.8303; SE = 1.63; p-value = 0.0823
- Q2 2017 – Q1 2018: estimate = 0.7693; SE = 1.52; p-value = 0.6120
- Q3 2017 – Q2 2018: estimate = 0.0375; SE = 1.59; p-value = 0.9812

Other

- Q3 2016 – Q2 2017: estimate = -1.2128; SE = 0.28; p-value < 0.0001
- Q4 2016 – Q3 2017: estimate = -0.9606; SE = 0.27; p-value = 0.0003

- Q1 2017 – Q4 2017: estimate = -1.1116; SE = 0.26; p-value < 0.0001
- Q2 2017 – Q1 2018: estimate = -1.2595; SE = 0.26; p-value < 0.0001
- Q3 2017 – Q2 2018: estimate = -1.1814; SE = 0.25; p-value < 0.0001

Hispanic Ethnicity

- Q3 2016 – Q2 2017: estimate = 1.3465; SE = 0.42; p-value = 0.0012
- Q4 2016 – Q3 2017: estimate = 1.0871; SE = 0.41; p-value = 0.0081
- Q1 2017 – Q4 2017: estimate = 1.1657; SE = 0.40; p-value = 0.0036
- Q2 2017 – Q1 2018: estimate = 1.4814; SE = 0.40; p-value = 0.0002
- Q3 2017 – Q2 2018: estimate = 1.3907; SE = 0.38; p-value = 0.0003

Not Currently Married*

- Q3 2016 – Q2 2017: estimate = -0.2334; SE = 0.16; p-value = 0.1364
- Q4 2016 – Q3 2017: estimate = -0.2876; SE = 0.16; p-value = 0.0675
- Q1 2017 – Q4 2017: estimate = -0.1637; SE = 0.16; p-value = 0.2958
- Q2 2017 – Q1 2018: estimate = -0.2451; SE = 0.16; p-value = 0.1192
- Q3 2017 – Q2 2018: estimate = -0.1457; SE = 0.16; p-value = 0.3492

Payer Source (reference = Private Insurance)

Medicare Fee-For-Service

- Q3 2016 – Q2 2017: estimate = 0.0496; SE = 0.31; p-value = 0.8741
- Q4 2016 – Q3 2017: estimate = 0.3564; SE = 0.32; p-value = 0.2635
- Q1 2017 – Q4 2017: estimate = 0.2137; SE = 0.32; p-value = 0.4993
- Q2 2017 – Q1 2018: estimate = 0.3727; SE = 0.31; p-value = 0.2348
- Q3 2017 – Q2 2018: estimate = 0.5508; SE = 0.31; p-value = 0.0750

Medicare Managed Care

- Q3 2016 – Q2 2017: estimate = 0.3015; SE = 0.34; p-value = 0.3708
- Q4 2016 – Q3 2017: estimate = 0.6158; SE = 0.34; p-value = 0.0723
- Q1 2017 – Q4 2017: estimate = 0.1771; SE = 0.34; p-value = 0.5985
- Q2 2017 – Q1 2018: estimate = -0.0315; SE = 0.33; p-value = 0.9248
- Q3 2017 – Q2 2018: estimate = 0.0233; SE = 0.32; p-value = 0.9429

Medicaid Fee-For-Service

- Q3 2016 – Q2 2017: estimate = -0.4975; SE = 0.30; p-value = 0.0972
- Q4 2016 – Q3 2017: estimate = -0.1749; SE = 0.30; p-value = 0.5649
- Q1 2017 – Q4 2017: estimate = -0.2253; SE = 0.30; p-value = 0.4525
- Q2 2017 – Q1 2018: estimate = -0.2186; SE = 0.30; p-value = 0.4626
- Q3 2017 – Q2 2018: estimate = -0.2220; SE = 0.29; p-value = 0.4473

Medicaid Managed Care

- Q3 2016 – Q2 2017: estimate = -1.1333; SE = 0.36; p-value = 0.0016
- Q4 2016 – Q3 2017: estimate = -0.8581; SE = 0.36; p-value = 0.0182
- Q1 2017 – Q4 2017: estimate = -0.8984; SE = 0.36; p-value = 0.0123
- Q2 2017 – Q1 2018: estimate = -0.4993; SE = 0.35; p-value = 0.1556

- Q3 2017 – Q2 2018: estimate = -0.2979; SE = 0.34; p-value = 0.3854

Other Payer Source**

- Q3 2016 – Q2 2017: estimate = 0.3060; SE = 0.28; p-value = 0.2775
- Q4 2016 – Q3 2017: estimate = 0.5650; SE = 0.29; p-value = 0.0489
- Q1 2017 – Q4 2017: estimate = 0.4783; SE = 0.28; p-value = 0.0922
- Q2 2017 – Q1 2018: estimate = 0.2834; SE = 0.28; p-value = 0.3120
- Q3 2017 – Q2 2018: estimate = 0.3872; SE = 0.28; p-value = 0.1593

Unknown

- Q3 2016 – Q2 2017: estimate = -0.6519; SE = 1.10; p-value = 0.5522
- Q4 2016 – Q3 2017: estimate = -0.3878; SE = 1.04; p-value = 0.7079
- Q1 2017 – Q4 2017: estimate = -0.4957; SE = 1.06; p-value = 0.6393
- Q2 2017 – Q1 2018: estimate = -0.7632; SE = 1.03; p-value = 0.4571
- Q3 2017 – Q2 2018: estimate = -1.2370; SE = 0.97; p-value = 0.2016

* Includes never married, widowed, separated, divorced, and not assessed/no information.

** Includes Private insurance/Medigap, Workers' compensation, title programs, other government sources, self-pay, or no

payor source.

Note: SE=Standard error; Patient-level exclusion criteria applied.

Source: RTI analysis of LTCH CARE Data Set, July 2016 – June 2018. (Program reference: 2632_03)

1b.5. If no or limited data on disparities from the measure as specified is reported in 1b.4, then provide a summary of data from the literature that addresses disparities in care on the specific focus of measurement. Include citations. Not necessary if performance data provided in 1b.4

We were unable to identify literature that reported disparities data (race/ethnicity, gender, age, insurance status, socioeconomic status or disability status) related to functional improvement among LTCH patients.

2. Reliability and Validity—Scientific Acceptability of Measure Properties

Extent to which the measure, as specified, produces consistent (reliable) and credible (valid) results about the quality of care when implemented. ***Measures must be judged to meet the sub criteria for both reliability and validity to pass this criterion and be evaluated against the remaining criteria.***

2a.1. Specifications The measure is well defined and precisely specified so it can be implemented consistently within and across organizations and allows for comparability. eMeasures should be specified in the Health Quality Measures Format (HQMF) and the Quality Data Model (QDM).

De.5. Subject/Topic Area *(check all the areas that apply):*

Cardiovascular, Cardiovascular : Congestive Heart Failure, Critical Care, Respiratory : Pneumonia

De.6. Non-Condition Specific*(check all the areas that apply):*

Health and Functional Status : Change

De.7. Target Population Category *(Check all the populations for which the measure is specified and tested if any):*

Populations at Risk : Individuals with multiple chronic conditions

S.1. Measure-specific Web Page *(Provide a URL link to a web page specific for this measure that contains current detailed specifications including code lists, risk model details, and supplemental materials. Do not enter a URL linking to a home page or to general information.)*

https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/Post-Acute-Care-Quality-Initiatives/Functional-Measures-.html

S.2a. If this is an eMeasure, HQMF specifications must be attached. Attach the zipped output from the eMeasure authoring tool (MAT) - if the MAT was not used, contact staff. (Use the specification fields in this online form for the plain-language description of the specifications)

This is not an eMeasure **Attachment:**

S.2b. Data Dictionary, Code Table, or Value Sets *(and risk model codes and coefficients when applicable) must be attached. (Excel or csv file in the suggested format preferred - if not, contact staff)*

Attachment **Attachment:** Change_in_Mobility_NQF_2632_Risk_Adj_Model_01-07-2019-636824735650484277.xlsx

S.2c. Is this an instrument-based measure (i.e., data collected via instruments, surveys, tools, questionnaires, scales, etc.)? Attach copy of instrument if available.

Attachment **Attachment:** LTCH_CARE_Data_Set_Version_4.00_-_Admission_and_Planned_Discharge_combined.pdf

S.2d. Is this an instrument-based measure (i.e., data collected via instruments, surveys, tools, questionnaires, scales, etc.)? Attach copy of instrument if available.

Clinician

S.3.1. For maintenance of endorsement: Are there changes to the specifications since the last updates/submission. If yes, update the specifications for S1-2 and S4-22 and explain reasons for the changes in S3.2.

Yes

S.3.2. For maintenance of endorsement, please briefly describe any important changes to the measure specifications since last measure update and explain the reasons.

We have made updates to the risk adjustors since last endorsement:

(1) Risk-Adjustors: We have updated the risk adjustment model by removing several comorbidities. Rationale: When examining the risk adjustment model using the national LTCH data, we found some comorbidities were no longer significant predictors of change in mobility or the association between the comorbidity and functional outcomes was no longer consistent with the evidence from the literature or clinical expectations.

(2) Measure Calculation: The risk-adjustment procedure for this measure involves comparing patients' observed change in mobility scores with their expected change in mobility scores. We are revising this part of the measure calculation. The prior approach used the ratio of the observed to expected values and the ratio was multiplied by the national mean. The new approach uses the difference between the observed and expected values, and the difference value is added to the national mean. Rationale: We have developed a change in mobility performance measure for skilled nursing facilities (SNFs) and use the difference approach for the SNF measure given the potential for more variation in the observed and expected values due to a more heterogeneous SNF population. We examined the LTCH change in mobility performance measure data and noted some LTCHs had low expected change in mobility scores, which can lead to large ratio values. We are now updating this LTCH functional outcome measure to use the difference approach.

S.4. Numerator Statement *(Brief, narrative description of the measure focus or what is being measured about the target population, i.e., cases from the target population with the target process, condition, event, or outcome) DO NOT include the rationale for the measure.*

IF an OUTCOME MEASURE, state the outcome being measured. Calculation of the risk-adjusted outcome should be described in the calculation algorithm (S.14).

The measure does not have a simple form for the numerator and denominator. This measure estimates the risk-adjusted change in mobility score between admission and discharge among LTCH patients requiring ventilator support at admission. The change in mobility score is calculated as the difference between the discharge mobility score and the admission mobility score.

S.5. Numerator Details *(All information required to identify and calculate the cases from the target population with the target process, condition, event, or outcome such as definitions, time period for data collection, specific data collection items/responses, code/value sets – Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format at S.2b)*

IF an OUTCOME MEASURE, describe how the observed outcome is identified/counted. Calculation of the risk-adjusted outcome should be described in the calculation algorithm (S.14).

Eight mobility activities (listed below) are each scored by a clinician based on a patient´s ability to complete the activity. The scores for the 8 mobility activities are summed to obtain a mobility score at the time of admission and discharge. The change in mobility is the difference between the discharge mobility score and the admission mobility score.

The 8 mobility items are:

GG0170A. Roll left and right

GG0170B. Sit to lying

GG0170C. Lying to sitting on side of bed

GG0170D. Sit to stand

GG0170E. Chair/bed-to-chair transfer

GG0170F. Toilet transfer

GG0170J. Walk 50 feet with two turns

GG0170K. Walk 150 feet

Each patient´s ability to complete each mobility activity (item) is rated by a clinician using the following 6-level rating scale:

level 06 - Independent

level 05 - Setup or clean up assistance

level 04 - Supervision or touching assistance

level 03 - Partial/moderate assistance

level 02 - Substantial/maximal assistance

level 01 - Dependent

If the patient did not attempt the activity, the reason that the activity did not occur is reported as:

07 = Patient refused

09 = Not applicable

10 = Not attempted due to environmental limitations

88 = Not attempted due to medical condition or safety concerns.

The performance period is 24 months for reporting on CMS's LTCH Compare website.

S.6. Denominator Statement *(Brief, narrative description of the target population being measured)*

The target population (denominator) for this quality measure is the number of LTCH patients requiring ventilator support at the time of admission to the LTCH.

S.7. Denominator Details *(All information required to identify and calculate the target population/denominator such as definitions, time period for data collection, specific data collection items/responses, code/value sets – Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format at S.2b.)*

IF an OUTCOME MEASURE, describe how the target population is identified. Calculation of the risk-adjusted outcome should be described in the calculation algorithm (S.14).

The denominator includes all LTCH patients requiring ventilator support on admission who are discharged during the performance period, including patients age 21 and older with all payer sources. Patients are selected based on submitted LTCH Care Data Set Admission and Discharge assessment forms.

S.8. Denominator Exclusions *(Brief narrative description of exclusions from the target population)*

This quality measure has following patient-level exclusion criteria:

1) Patients with incomplete stays:

Rationale: When a patient has an incomplete stay, for example, the patients leave urgently due to a medical emergency, it can be challenging to gather accurate discharge functional status data. Patients with incomplete stays include patients who are unexpectedly discharged to an acute-care setting (Inpatient Prospective Payment System or Inpatient Psychiatric Hospital or unit) because of a medical emergency or psychiatric condition; patients transferred to another LTCH; patients who leave the LTCH against medical advice; patients who die; and patients with a length of stay less than 3 days.

2) Patients discharged to hospice:

Rationale: Patients discharged to hospice are excluded because functional improvement may not be a goal for these patients.

3) Patients with progressive neurological conditions, including amyotrophic lateral sclerosis, multiple sclerosis, Parkinson's disease, and Huntington's chorea:

Rationale: These patients are excluded because they may have functional decline or less predictable function trajectories.

4) Patients in coma, persistent vegetative state, complete tetraplegia, and locked-in syndrome:

Rationale: The patients are excluded because they may have limited or less predictable mobility recovery.

5) Patients younger than age 21:

Rationale: There is only limited evidence published about functional outcomes for individuals younger than 21.

6) Patients who are coded as independent on all the mobility items at admission:

Rationale: These patients are excluded because no improvement in mobility skills can be measured with the mobility items used in this quality measure.

Facility-level quality measure exclusion: For LTCHs with fewer than 20 patient stays, data for this quality measure are not publicly reported.

S.9. Denominator Exclusion Details *(All information required to identify and calculate exclusions from the denominator such as definitions, time period for data collection, specific data collection items/responses, code/value sets – Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format at S.2b.)*

For each of the following exclusion criteria, we provide the data collection items used to identify patient records to be excluded. These items are on the LTCH CARE Data Set Version 4.00.

1) Patients with incomplete stays include patients who are unexpectedly discharged to an acute-care setting (Inpatient Prospective Payment System or Inpatient Psychiatric Hospital or unit) because of a medical emergency or psychiatric condition; patients transferred to another LTCH; patients who leave the LTCH against medical advice; patients who die; and patients with a length of stay less than 3 days.

Items used to identify these patient records:

A2110. Discharge Location

04 = Hospital emergency department

05 = Short-stay acute hospital (IPPS)

06 = Long-term care hospital (LTCH)

08 = Psychiatric hospital or unit

12 = Discharged Against Medical Advice

A0250. Reason for Assessment

11 = Unplanned discharge

12 = Expired

Patients with a length of stay less than 3 days:

We calculate length of stay using the following items on the LTCH CARE Data Set.

A0220. Admission Date

A0270. Discharge Date

Length of stay is calculated as the Discharge Date minus the Admission Date (Discharge Date - Admission Date). Patient records with a length of stay less than 3 days are excluded.

2) Patients discharged to hospice

Items used to identify these patient records:

A2110. Discharge Location

10 = Hospice

3) Patients with progressive neurological conditions, including amyotrophic lateral sclerosis, multiple sclerosis, Parkinson's disease, and Huntington's chorea are excluded because these patients may have less predictable mobility recovery or functional decline may be expected.

Items used to identify these patient records:

I5450. Amyotrophic Lateral Sclerosis = 1

I5200. Multiple Sclerosis = 1, or

I5300. Parkinson's Disease = 1, or

I5250. Huntington´s Disease = 1.

4) Patients in coma, persistent vegetative state, severe anoxic brain damage, cerebral edema, or compression of brain, complete tetraplegia, and locked-in syndrome are excluded, because they may have limited or less predictable mobility recovery.

Items used to identify these patient records:

B0100. Comatose = 1, or;

I5101. Complete Tetraplegia = 1, or;

I5460. Locked-In State = 1, or;

I5470. Severe Anoxic Brain Damage, Cerebral Edema, or Compression of Brain.

5) Patients younger than 21 at the time of admission

Items used to identify these patient records:

A0900. Birth Date

A0220. Admission Date

6) Patients who are coded as independent (score = 06) on all the mobility items at admission

Items used to identify these patient records at admission:

GG0170A. Roll left and right = 06, and;

GG0170B. Sit to lying = 06, and;

GG0170C. Lying to sitting on side of bed = 06, and;

GG0170D. Sit to stand, = 06 and,

GG0170E. Chair/bed-to-chair transfer, = 06, and;

GG0170F. Toilet transfer, = 06, and;

GG0170J. Walk 50 feet with two turns = 06, and;

GG0170K. Walk 150 feet = 06.

S.10. Stratification Information *(Provide all information required to stratify the measure results, if necessary, including the stratification variables, definitions, specific data collection items/responses, code/value sets, and the risk-model covariates and coefficients for the clinically-adjusted version of the measure when appropriate – Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format with at S.2b.)*

Not applicable. This measure does not use stratification for risk-adjustment.

S.11. Risk Adjustment Type (Select type. Provide specifications for risk stratification in measure testing attachment)

Statistical risk model

If other:

S.12. Type of score:

Continuous variable, e.g. average

If other:

S.13. Interpretation of Score *(Classifies interpretation of score according to whether better quality is associated with a higher score, a lower score, a score falling within a defined interval, or a passing score)*

Better quality = Higher score

S.14. Calculation Algorithm/Measure Logic (*Diagram or describe the calculation of the measure score as an ordered sequence of steps including identifying the target population; exclusions; cases meeting the target process, condition, event, or outcome; time period for data, aggregating data; risk adjustment; etc.)*

We provide the detailed calculation algorithm in an attachment entitled "LTCH Detailed Function QM Specifications 2632 01-07-2019" included in the Appendix.

The detailed calculation algorithm is provided to the public in the document entitled LTCH Measure Calculations and Reporting User's Manual. The current version of this document is available at: https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/LTCH-Quality-Reporting/LTCH-Quality-Reporting-Measures-Information.html.

The following are the key steps used to calculate the measure:

1) Sum the scores of the admission mobility items to create an admission mobility score for each patient. Mobility items that contained 'activity not attempted' codes (07. Patient refused, 09. Not applicable, 10. Not attempted due to environmental limitations, and 88. Not attempted due to medical condition or safety concerns) or were skipped, dashed, or missing are recoded to 01. Dependent (range: 8 to 48).

2) Sum the scores of the discharge mobility items to create a discharge mobility score for each patient. Mobility items that contained 'activity not attempted' values (07. Patient refused, 09. Not applicable, 10. Not

attempted due to environmental limitations, and 88. Not attempted due to medical condition or safety concerns) or were skipped, dashed, or missing are recoded to 01. Dependent (range: 8 to 48).

3) Identify the records of patients who meet the exclusion criteria and exclude these patient records from analyses.

4) Calculate the difference between the admission mobility score (from step 1) and the discharge mobility score (from step 2) for each patient to create a change in mobility score for each patient.

5) Calculate an expected change in mobility score for each patient using regression coefficients from national data and each patient's admission characteristics (risk adjustors).

6) Calculate an average observed change in mobility score for each LTCH (using the patient data calculated in step 4). This is the facility-level observed change in mobility score.

7) Calculate an average expected change in mobility score for each LTCH (using the patient data calculated in step 5). This is the facility-level expected change in mobility score.

8) Subtract the facility-level expected change score from the facility-level observed change score to determine the difference in scores (difference value). A difference value that is 0 indicates the observed and expected scores are equal. An observed minus expected difference value that is higher than 0 (positive) indicates that the observed change score is greater (better) than the expected change score. An observed minus expected difference value that is less than 0 (negative) indicates that the observed change score is lower (worse) than the expected change score.

9) Add the national average change in mobility score to each LTCH's difference value (from step 8). This is the risk-adjusted mean change in mobility score.

Each patient´s ability to complete each mobility activity (item) is rated by a clinician using the following 6-level rating scale:

level 06 - Independent

level 05 - Setup or clean up assistance

level 04 - Supervision or touching assistance

level 03 - Partial/moderate assistance

level 02 - Substantial/maximal assistance

level 01 - Dependent

The 8 mobility items are:

GG0170A. Roll left and right

GG0170B. Sit to lying

GG0170C. Lying to sitting on side of bed

GG0170D. Sit to stand

GG0170E. Chair/bed-to-chair transfer

GG0170F. Toilet transfer

GG0170J. Walk 50 feet with two turns

GG0170K. Walk 150 feet

S.15. Sampling *(If measure is based on a sample, provide instructions for obtaining the sample and guidance on minimum sample size.)*

IF an instrument-based performance measure (e.g., PRO-PM), identify whether (and how) proxy responses are allowed.

Not applicable. This measure uses LTCH CARE Data Set data for all Medicare patients treated by LTCHs for the performance period. There is no sampling. This is an instrument-based measure that relies on clinician-reported data, therefore proxy responses are not relevant.

S.16. Survey/Patient-reported data *(If measure is based on a survey or instrument, provide instructions for data collection and guidance on minimum response rate.)*

Specify calculation of response rates to be reported with performance measure results.

Not applicable. This measure uses clinician-reported data.

S.17. Data Source *(Check ONLY the sources for which the measure is SPECIFIED AND TESTED).*

If other, please describe in S.18.

Instrument-Based Data

S.18. Data Source or Collection Instrument *(Identify the specific data source/data collection instrument (e.g. name of database, clinical registry, collection instrument, etc., and describe how data are collected.)*

IF instrument-based, identify the specific instrument(s) and standard methods, modes, and languages of administration.

LTCH CARE Data Set

S.19. Data Source or Collection Instrument *(available at measure-specific Web page URL identified in S.1 OR in attached appendix at A.1)*

Available at measure-specific web page URL identified in S.1

S.20. Level of Analysis *(Check ONLY the levels of analysis for which the measure is SPECIFIED AND TESTED)*

Facility

S.21. Care Setting *(Check ONLY the settings for which the measure is SPECIFIED AND TESTED)*

Post-Acute Care

If other:

S.22. COMPOSITE Performance Measure - Additional Specifications *(Use this section as needed for aggregation and weighting rules, or calculation of individual performance measures if not individually endorsed.)*

Not applicable. This is not a composite measure.

2. Validity – See attached Measure Testing Submission Form

NQF_LTCH_Mobility_Testing_Final.docx,2632_nqf_testing_4-22-2019.docx

2.1 For maintenance of endorsement

Reliability testing: If testing of reliability of the measure score was not presented in prior submission(s), has reliability testing of the measure score been conducted? If yes, please provide results in the Testing attachment. Please use the most current version of the testing attachment (v7.1). Include information on all testing conducted (prior testing as well as any new testing); use red font to indicate updated testing.

Yes

2.2 For maintenance of endorsement

Has additional empirical validity testing of the measure score been conducted? If yes, please provide results in the Testing attachment. Please use the most current version of the testing attachment (v7.1). Include information on all testing conducted (prior testing as well as any new testing); use red font to indicate updated testing.

Yes

2.3 For maintenance of endorsement

Risk adjustment: For outcome, resource use, cost, and some process measures, risk-adjustment that includes social risk factors is not prohibited at present. Please update sections 1.8, 2a2, 2b1,2b4.3 and 2b5 in the Testing attachment and S.140 and S.11 in the online submission form. NOTE: These sections must be updated even if social risk factors are not included in the risk-adjustment strategy. You MUST use the most current version of the Testing Attachment (v7.1) -- older versions of the form will not have all required questions.

Yes - Updated information is included

Measure Testing (subcriteria 2a2, 2b1-2b6)

Measure Number (*if previously endorsed*)**:** 2632
Measure Title: Long-Term Care Hospital (LTCH) Functional Outcome Measure: Change in Mobility Among Patients Requiring Ventilator Support
Date of Submission: 1/7/2019

Type of Measure:

☒ **Outcome (*including PRO-PM*)**	☐ **Composite – *STOP – use composite testing form***
☐ Intermediate Clinical Outcome	☐ Cost/resource
☐ Process *(including Appropriate Use)*	☐ Efficiency
☐ Structure	

1. DATA/SAMPLE USED FOR ALL TESTING OF THIS MEASURE

Often the same data are used for all aspects of measure testing. In an effort to eliminate duplication, the first five questions apply to all measure testing. If there are differences by aspect of testing,(e.g., reliability vs. validity) be sure to indicate the specific differences in question 1.7.

1.1. What type of data was used for testing? (*Check all the sources of data identified in the measure specifications and data used for testing the measure. Testing must be provided for all the sources of data specified and intended for measure implementation.* ***If different data sources are used for the numerator and denominator, indicate N [numerator] or D [denominator] after the checkbox.***)

Measure Specified to Use Data From: (*must be consistent with data sources entered in S.17*)	**Measure Tested with Data From:**
☐ abstracted from paper record	☐ abstracted from paper record
☐ claims	☐ claims
☐ registry	☐ registry
☐ abstracted from electronic health record	☐ abstracted from electronic health record
☐ eMeasure (HQMF) implemented in EHRs	☐ eMeasure (HQMF) implemented in EHRs
☒ other: Long-Term Care Hospital (LTCH) Continuity Assessment Record and Evaluation (CARE) Data Set	☒ other: Long-Term Care Hospital (LTCH) Continuity Assessment Record and Evaluation (CARE) Data Set

1.2. If an existing dataset was used, identify the specific dataset (*the dataset used for testing must be consistent with the measure specifications for target population and healthcare entities being measured; e.g., Medicare Part A claims, Medicaid claims, other commercial insurance, nursing home MDS, home health OASIS, clinical registry*).

The primary dataset used for calculating this performance measure was the national LTCH CARE Data Set data. A copy of the LTCH CARE Data Set can be found on the following website:

https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/LTCH-Quality-Reporting/LTCH-CARE-Data-Set-and-LTCH-QRP-Manual.html

We used one additional data source for measure testing only to provide facility and patient-level characteristics not available in the LTCH CARE Data Set. This source is not used for quality measure calculation:

For reliability and validity testing analyses that involved facility characteristics, we used the Provider of Service file.

- **Provider of Services Current Files (POS File):** We used the POS file to describe the characteristics of LTCHs, such as census region, ownership type, and rurality, reported in **Table 1**. The POS file contains data on characteristics of hospitals and other types of healthcare facilities, including the name and address of the facility and the type of Medicare services the facility provides, among other information. The data are collected through the CMS Regional Offices. General information about the POS Files is available at: https://www.cms.gov/Research-Statistics-Data-and-Systems/Downloadable-Public-Use-Files/Provider-of-Services/index.html.

1.3. What are the dates of the data used in testing?

For most testing reported in this document, we analyzed patients discharged from July 1, 2016 through March 31, 2018 (21 Months). While this performance measure has a 24-month performance period, we only have access to complete and finalized data for 21 months at this time. This is because national data collection for this performance measure began on April 1, 2016, and complete admission and discharge for all LTCH patients has only been available since July 1, 2016. We anticipate having updated testing results using 24 months of data by June 2019. For the Rasch analysis, we analyzed patients discharged in fiscal year 2017 (October 1, 2016 through September 30, 2017; 12 Months)

1.4. What levels of analysis were tested? (*testing must be provided for all the levels specified and intended for measure implementation, e.g., individual clinician, hospital, health plan*)

Measure Specified to Measure Performance of: (*must be consistent with levels entered in item S.20*)	**Measure Tested at Level of:**
☐ individual clinician	☐ individual clinician
☐ group/practice	☐ group/practice
☒ hospital/facility/agency	☒ hospital/facility/agency
☐ health plan	☐ health plan
☐ other:	☐ other:

1.5. How many and which measured entities were included in the testing and analysis (by level of analysis and data source)? (*identify the number and descriptive characteristics of measured entities included in the analysis (e.g., size, location, type); if a sample was used, describe how entities were selected for inclusion in the sample*)

Long-Term Care Hospitals Included in the LTCH CARE Data Set

Testing for this performance measure involved several types of data element, scale/instrument and computed performance measure score reliability and validity analyses, performance measure score variability analyses, and social risk factor analysis. The unit of analysis for the data element and scale/instrument analyses is patient assessments or patient stays, and the unit of analysis for the computed performance measure score analyses is providers (i.e., LTCHs). National data collection for the change in mobility among patients requiring ventilator support began April 1, 2016 with the release of LTCH CARE Data Set Version 3.00.

A total of 413 LTCHs submitted LTCH CARE Data Set assessments for patients discharged during the testing period, July 2016 – March 2018.

Table 1 displays the geographical location and provider characteristics of LTCHs that reported LTCH CARE Data Set data for this performance measure. The majority of these LTCHs are located in the southern (CMS Regions 4 and 6) and midwestern states (CMS Region 5) with nearly 30 percent in Region 6 (TX, LA, AR, OK, NM). The majority of LTCHs are in urban settings (95.1%) and under private ownership (69.4%).

Table 1. Number of LTCHs Reporting by Facility Characteristics, July 2016 – March 2018 (N=413)

Characteristic	Number (Percent)
CMS Region	
Region 1: CT, ME, MA, NH, RI, VT	13 (3.2%)
Region 2: PR, VI, NY, NJ	9 (2.2%)
Region 3: MD, DC, DE, WV, VA, PA	30 (7.2%)
Region 4: NC, SC, TN, FL, GA, AL, KY, MS	92 (22.2%)
Region 5: MI, MN, OH, IL, IN, WI	71 (17.1%)
Region 6: TX, LA, AR, OK, NM	122 (29.5%)
Region 7: MO, KS, IA, NE	22 (5.3%)
Region 8: ND, UT, SD, WY, CO, MT	16 (3.8%)
Region 9: NV, AZ, CA, HI, AS, Pacific Territories	31 (7.5%)
Region 10: WA, AK, ID, OR	7 (1.6%)
Urbanicity	
Rural	20 (4.8%)
Urban	393 (95.1%)
Ownership Type	
Government	15 (3.6%)
Private	287 (69.4%)
Non-profit	98 (23.7%)
Other	13 (3.1%)

Note: Values are reported as frequency (percent)
Source: RTI analysis of LTCH CARE Data Set July 2016 – March 2018, and Provider of Service (POS) File (Program reference: 2632_01)

Rasch Analysis Sample using the LTCH CARE Data Set – Fiscal Year 2017 Data

As noted above, the reliability and validity testing that involved Rasch analysis and internal consistency was conducted using fiscal year 2017 data. This dataset included 403 LTCHs. The characteristics of these LTCHs are very similar to the provider data for the data reported above (July 2016 – March 2018).

1.6. How many and which patients were included in the testing and analysis (by level of analysis and data source)? (*identify the number and descriptive characteristics of patients included in the analysis (e.g., age, sex, race, diagnosis); if a sample was used, describe how patients were selected for inclusion in the sample*)

Total Number of Patients Included in the LTCH CARE Data Set – July 1, 2016 to March 31, 2018 Data

During the testing time period (July 1, 2016 through March 31, 2018), LTCHs submitted a total of 132,274 patient assessments (admission and discharge) representing 66,137 patient stays for patients requiring invasive ventilator support at admission. The sociodemographic characteristics of these patients are summarized in **Table 2**.

Patients older than the age of 65 accounted for 55.8 percent of these LTCH patients. Male patients comprised 54.7 percent, 64.1 percent of patients were white, and just over a quarter of patients were currently married. More than half of patients have either Medicare or Medicaid. The majority of patient stays ended with the patient discharged to another post-acute care setting (48.0%) or returning to short-term acute care hospital

(18.4%). Few patients were discharged to home with or without care from a home health service organization (10.9%).

Table 2. LTCH Patient Characteristics, July 2016 – March 2018 (N=66,137)

Characteristic	Number (Percent)
Age	
64 and younger	29,247 (44.2%)
65 to 69	10,932 (16.5%)
70 to 74	9,804 (14.8%)
75 to 79	7,847 (11.9%)
80 to 84	4,702 (7.1%)
85 and older	3,562 (5.4%)
Gender	
Male	36,177 (54.7%)
Female	29,960 (45.3%)
Race/Ethnicity*	
White	42,453 (64.1%)
Black or African American	13,243 (20.0%)
Asian	1,319 (1.9%)
American Indian/Alaskan Native	401 (0.6%)
Native Hawaiian/Pacific Islander	163 (0.2%)
None of the Above	8,558 (12.9%)
Hispanic or Latino	3,682 (5.5%)
Marital Status	
Married	18,591 (28.1%)
Not Currently Married**	47,546 (71.9%)
Payer Information***	
Medicare Fee-for-Service	10,363 (15.7%)
Medicare Managed Care	7,475 (11.3%)
Medicaid Fee-for-Service	14,228 (21.5%)
Medicaid Managed Care	6,151 (9.3%)
Private Managed Care	6,178 (9.3%)
Other****	21,391 (32.3%)
Unknown	351 (0.5%)
Discharge to Location	
Short-Term Acute Care General Hospital	12,197 (18.4%)
Home (with or without home care)	7,223 (10.9%)
Institutional Post-Acute Care[#]	31,736 (48.0%)
Other[†]	2,244 (3.4%)
Expired	11,832 (17.9%)
Discharged Against Medical Advice	192 (0.3%)
Not Listed	713 (1.1%)

Note: Values are reported as frequency (percent)

* Percentages can add up to more than 100%; if more than 1 category was selected the patient is assigned to both categories.

** Includes never married, widowed, separated, divorced, and not assessed/no information.
*** More than 1 payer source can be selected.
**** Includes Private insurance/Medigap, Workers' compensation, title programs, other government sources, self-pay, or no payor source.
[#] Includes institutional settings: skilled nursing facilities, inpatient rehabilitation facilities or another LTCH.
[†] Includes nursing homes, hospice, inpatient psychiatric facilities, and other intermediate care settings.
Source: RTI analysis of LTCH CARE Data Set, July 2016 – March 2018 (Program reference: 2632_01).

Rasch Analysis Sample using LTCH CARE Data Set – Fiscal Year 2017 Data

As noted above, the reliability and validity testing that involved Rasch analysis and internal consistency testing was conducted using fiscal year 2017 data. 322,963 randomly selected assessments from the LTCH CARE Data Set in fiscal year 2017 were analyzed for the fit assessment and internal consistency. The characteristics of these patients was very similar to LTCH patients discharged in fiscal year 2017.

1.7. If there are differences in the data or sample used for different aspects of testing (e.g., reliability, validity, exclusions, risk adjustment), identify how the data or sample are different for each aspect of testing reported below.

Most testing was conducted using LTCH CARE Data Set data submitted by LTCHs for patients discharged from July 1, 2016 through March 31, 2018 (**Tables 1** and **2**).

For the Rasch analyses and internal consistency analyses, we used a random subsample of assessments in the national data (n = 322,963) for patients discharged in fiscal year 2017. The Rasch analysis and internal consistency work include:

- Scale Construct Validity Testing - Item Difficulty Ordering
- Scale Validity Testing - Fit Assessment
- Data element Validity Testing - Response Option Assessment

1.8 What were the social risk factors that were available and analyzed? For example, patient-reported data (e.g., income, education, language), proxy variables when social risk data are not collected from each patient (e.g. census tract), or patient community characteristics (e.g. percent vacant housing, crime rate) which do not have to be a proxy for patient-level data.

We examined whether 4 social risk factors affected performance measure scores. The social risk factors are: 1) payer source (patient-level variable); 2) marriage status (patient-level variable); 3) race (patient-level variable); and 4) ethnicity (patient-level).

We selected these patient-level social risk factors based on our review of the literature showing functional outcomes can vary by payer source, living situation and race/ethnicity.

Payer source, marriage status, race, and ethnicity data were derived from the LTCH CARE Data Set.

2a2. RELIABILITY TESTING

***Note**: If accuracy/correctness (validity) of data elements was empirically tested, separate reliability testing of data elements is not required – in 2a2.1 check critical data elements; in 2a2.2 enter "see section 2b2 for validity testing of data elements"; and skip 2a2.3 and 2a2.4.*

2a2.1. What level of reliability testing was conducted? (*may be one or both levels*)
☒ **Critical data elements used in the measure** (*e.g., inter-abstractor reliability; data element reliability must address ALL critical data elements*)
☒ **Performance measure score** (e.g., *signal-to-noise analysis*)

2a2.2. For each level checked above, describe the method of reliability testing and what it tests (*describe the steps—do not just name a method; what type of error does it test; what statistical analysis was used*)

We report testing results throughout this document for data elements, the mobility scale/instrument and the computed performance measure score. To assist the reader in understanding the testing analysis and results, we begin by providing a brief overview of these components:

1. Mobility Data Elements:
    a. There are 8 mobility data elements, which are included in LTCH CARE Data Set Section GG. Depending on the context, we sometimes refer to these data elements as "items" or "activities."
    b. The mobility data are collected at the time of admission and discharge using a 6-level rating scale (01 to 06), or activity not attempted codes if, for example, the activity was not attempted due to medical or safety concerns.
    c. Higher scores indicate higher ability (i.e., more independence)
    d. For the performance measure calculation, data element activity not attempted codes and missing data are recoded.
2. Admission and Discharge Mobility Scores (Scale/Instrument)
    a. An admission mobility scale score is created by summing the 8 mobility data element scores, after re-coding. The admission mobility score can range from 8 to 48 mobility units.
    b. A discharge mobility scale score is created by summing the 8 data element scores, after re-coding. The range of the discharge mobility score is 8 to 48 mobility units.
    c. For the Admission and Discharge Mobility Scores, a score of 8 indicates the patient is dependent on a helper to perform all 8 mobility activities (i.e., data elements) and a score of 48 means the patient is independent on all 8 activities.
3. Observed Change in Mobility
    a. An observed change in Mobility score is calculated by subtracting the observed (unadjusted) Discharge Mobility Score from the observed (unadjusted) Admission Mobility Score.
    b. The potential range of the Observed Change in Mobility Scores is -40 to + 40. Most patients are expected to have improved mobility abilities, and thus we observe mostly positive values.
4. Calculated Performance Measure Score: Risk-Adjusted Change in Mobility Score
    a. The calculated performance measure score is a risk-adjusted Change in Mobility Score. The risk-adjustment project is described in S.14. Calculation Algorithm/Measure Logic on the NQF Intent to Submit form and the attached file "LTCH_Detailed_Function_QM_Specifications_2632_01-07-2019.docx."
    b. This performance measure does not have a simple form for the numerator and denominator. This performance measure estimates the risk-adjusted mean change in mobility score between admission and discharge for LTCH patients requiring ventilator support.

Computed Performance Measure Score Reliability – Split-half Reliability (unit of analysis is providers): Split-half reliability was used to examine the reliability of the computed performance measure scores. The computed performance measure score is the risk-adjusted change in mobility score. For LTCHs with fewer than 20 patient stays, computed performance measure scores are not displayed to the public, therefore, we included facilities with 20 or more stays in this analysis. We conducted split-half reliability by randomly splitting each provider's patient stays into two groups and calculating correlations between the computed performance measure scores of the randomly divided groups. When a provider's data, after being randomly divided into two groups, show similar scores to one another, the performance measure score is more likely to reflect systematic differences in LTCH provider quality rather than random variation. The Pearson Product-Moment Correlation (r), Spearman Rank Correlation (ρ), and Intraclass Correlation Coefficient (ICC) were used to examine the performance measure reliability. Intraclass correlations were also calculated by facility volume quartile to examine whether there were differences in performance measure reliability by LTCH size.

Mobility Scale/Instrument Analysis- Internal Consistency (unit of analysis is patient assessments): In addition to the provider-level reliability testing of the computed performance measure scores described above, we examined the internal consistency of the mobility scale/instrument scores for each patient stay. Internal consistency provides a general assessment of how well the mobility items interrelate within the mobility

scale/instrument. This internal consistency analysis is an indicator of the reliability of the mobility scale/instrument and is thus a test of the reliability of the data elements.

Internal consistency was assessed using the Cronbach's alpha coefficient, which is the average correlation of all possible half-scale divisions. Cronbach's alpha is a statistic frequently calculated when testing instrument or scale psychometrics. The Cronbach's alpha reliability estimate ranges from zero to one, with an estimate of zero indicating that there is no consistency of measurement among the items, and one indicating perfect consistency. Many cutoff criteria exist to determine whether or not a scale shows good consistency or whether the items "hang together" well. Nunnally (1978) indicated that Cronbach's alpha should be at least 0.90 for item sets used in decision making. The internal consistency from the Rasch analysis assesses items using the KR20 (a special case of Cronbach's alpha) estimate, with the same cut-off requirements.

Citation: Nunnally, J. (1978). *Psychometric methods*. New York, NY: McGraw-Hill.

Critical Data Elements Testing using CARE Tool Data (2014) – Inter-Rater Reliability, Video (Standardized Patient) Reliability and Validity Testing (unit of analysis is patients): In our 2014 NQF testing document, we described several types of data element and scale/instrument reliability and validity analysis using data collected by providers as part of the Post-Acute Care Payment Reform Demonstration (2007-2012). This reliability and validity testing included the mobility data elements, as well as data elements that are used as risk adjustors for this performance measure. For more information about the development and testing of the data elements and scale/instrument, please see:

- Gage BJ, Constantine R, Aggarwal MM, Bernard S, Munevar D, Garrity M, Deutsch A, et al. (June, 2012). *The Development of the Continuity Assessment Record and Evaluation (CARE) Tool: Final Report*. Prepared for the Centers for Medicare & Medicaid Services. Available at: https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/Post-Acute-Care-Quality-Initiatives/Downloads/The-Development-and-Testing-of-the-Continuity-Assessment-Record-and-Evaluation-CARE-Item-Set-Final-Report-on-the-Development-of-the-CARE-Item-Set-Volume-1-of-3.pdf
- Gage BJ, Smith LM, Ross J, Coots LA, Shamsuddin KM, Deutsch A, Mallinson T, Reilly KE, Abbate JH, Gage-Croll Z. (August, 2012). *The development and testing of the Continuity Assessment Record and Evaluation (CARE) Item Set: Final Report on Reliability Testing, Volume 2 of 3*. Prepared for Centers for Medicare & Medicaid Services. Available at: https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/Post-Acute-Care-Quality-Initiatives/Downloads/The-Development-and-Testing-of-the-Continuity-Assessment-Record-and-Evaluation-CARE-Item-Set-Final-Report-on-Reliability-Testing-Volume-2-of-3.pdf
- Gage BJ, Deutsch A, Smith LM, Schwartz C, Ross J, Coots LA, Reilly KE, Abbate JH, Shamsuddin KM, Silver BC, et al. (September, 2012). *The Development and Testing of the Continuity Assessment Record and Evaluation (CARE) Item Set: Final Report on CARE Item Set and Current Assessment Comparisons, Volume 3 of 3*. Prepared for Centers for Medicare & Medicaid Services. Available at: https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/Post-Acute-Care-Quality-Initiatives/Downloads/The-Development-and-Testing-of-the-Continuity-Assessment-Record-and-Evaluation-CARE-Item-Set-Final-Report-on-the-Development-of-the-CARE-Item-Set-and-Current-Assessment-Comparisons-Volume-3-of-3.pdf
- Smith LM, Deutsch A, Hand LB, Etlinger AL, Ross J, Abbate JH, Gage-Croll Z, Barch D, Gage BJ. (September, 2012). *Continuity Assessment Record and Evaluation (CARE) Item Set: Additional Provider-Type Specific Interrater Reliability Analyses*. Prepared for Centers for Medicare & Medicaid Services. Available at: https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/Post-Acute-Care-Quality-Initiatives/Downloads/Continuity-Assessment-Record-and-Evaluation-CARE-Item-Set-Additional-Provider-Type-Specific-Interrater-Reliability-Analyses.pdf
- Smith LM, Deutsch A, Barch D, Ross J, Shamsuddin KM, Abbate JH, Schwartz C, Gage BJ. (September, 2012). *Continuity Assessment Record and Evaluation (CARE) Item Set: Video Reliability Testing*.

Prepared for Centers for Medicare & Medicaid Services. Available at: https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/Post-Acute-Care-Quality-Initiatives/Downloads/Continuity-Assessment-Record-and-Evaluation-CARE-Item-Set-Video-Reliability-Testing.pdf

- Gage BJ, Morley MA, Smith LM, Ingber MJ, Deutsch A, Kline TL, Dever JA, Abbate JH, Miller RD, Lyda-McDonald B, Kelleher CA, Garfinkel DB, Manning JR, Murtaugh CM, Stineman MG, Mallinson T. (March, 2012). *Post-Acute Care Payment Reform Demonstration: Final Report Volumes 1-4*. Prepared for the Centers for Medicare and Medicaid Services. Available at: https://www.cms.gov/Research-Statistics-Data-and-Systems/Statistics-Trends-and-Reports/Reports/Research-Reports-Items/PAC_Payment_Reform_Demo_Final.html

For more information on the history of the development of this functional status performance measure, please visit CMS's Post-Acute Care Quality Initiatives Function Measures website: https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/Post-Acute-Care-Quality-Initiatives/Functional-Measures-.html

Summary of critical data element reliability testing:

The inter-rater reliability of the data elements was tested in a subset of 34 providers (acute hospitals, HHAs, IRFs, LTCHs, and SNFs) distributed across 11 geographic areas. Each provider completed a duplicate Item Set (admission or discharge assessment) on 10–20 patients. The overall sample size was 449 for mobility items (448 for transfers). The weighted kappa values for the mobility items ranged between 0.558 for walk 150 feet to 0.901 for sitting to standing and chair/bed to chair transfer. Unweighted kappas ranged from 0.667 for walk 10 feet to 0.762 for sit to stand. In summary, kappa statistics indicated substantial agreement of data element codes among raters.

The video reliability study indicated substantial agreement with the mode and clinical team for the lying-to-sitting, sit-to-stand, chair/bed to chair transfer, and toilet transfer items (greater than 76%). Although rates of agreement with the mode and clinical team response were generally identical, for the toilet transfer item, the clinical team agreement is slightly lower. The items for walking and wheeling distances showed more variable levels of agreement across disciplines, with overall agreement generally in the moderate range (50–78%). For the Walk 10 feet item, there was a notable decrease in the agreement with the clinical team compared to agreement with the mode. This occurred because in two of the four videos where this item was assessed, the clinical team response differed from the mode.

Please see Appendix B for additional details about the inter-rater reliability and video reliability testing.

2a2.3. For each level of testing checked above, what were the statistical results from reliability testing? (e.*g., percent agreement and kappa for the critical data elements; distribution of reliability statistics from a signal-to-noise analysis*)

Computed Performance Measure Score Reliability (Unit of analysis is provider): Split-half analysis results (**Table 3**) indicated positive moderate-to-strong correlations ($r = 0.714$, $\rho = 0.710$, ICC= 0.714, $p = <0.0001$) between the LTCH providers' randomly divided groups' computed performance measure scores on the Change in Mobility Among Patients Requiring Ventilator Support performance measure, providing evidence of measure reliability. ICCs remained moderate-to-strong when stratifying by provider volume quartile, with ICCs for the volume quartiles ranging from 0.600 (20 – 44 discharges) to 0.807 (119 – 547 discharges).

Table 3. Interclass Correlation Coefficient by LTCH Volume, July 2016 – March 2018 (N=343)

Volume Quartile	Number of LTCHs	ICC
Quartile 1: 20 – 44	89	0.600
Quartile 2: 45 - 76	86	0.704
Quartile 3: 77 - 118	83	0.733
Quartile 4: 119 - 547	85	0.807
Total	343	0.714

Note: Providers with < 20 stays during the 21-month testing period are excluded.
Source: RTI analysis of LTCH CARE Data Set July 2016 – March 2018 (Program reference: 2632_reliability)

Scale/instrument Reliability - Internal Consistency (unit of analysis is patient stays): Assessments of the mobility data elements showed excellent reliability statistics. The overall Cronbach's alpha is 0.92.

2a2.4 What is your interpretation of the results in terms of demonstrating reliability? (i.*e., what do the results mean and what are the norms for the test conducted?*)

The analysis of the 2016-2018 data show that provider-level reliability of the computed performance measure scores was moderate-to-strong overall and when stratified by provider volume. The patient-level analysis of fiscal year 2017 data of the scale/instrument reliability showed excellent reliability.

Critical data element inter-rater reliability and video reliability testing found substantial reliability overall.

2b1. VALIDITY TESTING

2b1.1. What level of validity testing was conducted? (*may be one or both levels*)
☒ **Critical data elements** (*data element validity must address ALL critical data elements*)

☒ **Performance measure score**

☒ **Empirical validity testing**

☐ **Systematic assessment of face validity of performance measure score as an indicator** of quality or resource use (*i.e., is an accurate reflection of performance on quality or resource use and can distinguish good from poor performance*) **NOTE**: Empirical validity testing is expected at time of maintenance review; if not possible, justification is required.

2b1.2. For each level of testing checked above, describe the method of validity testing and what it tests (*describe the steps—do not just name a method; what was tested, e.g., accuracy of data elements compared to authoritative source, relationship to another measure as expected; what statistical analysis was used)*

Scale/Instrument Content Validity - Similarity of Data Elements Across Other Mobility Assessment Instruments: Patient functioning is a construct that is often measured based on patient abilities, and the activities (data elements) included in functional assessment instruments vary. We compared the list of Section GG data elements used to calculate the Change in Mobility Among Patients Requiring Ventilator Support performance measure with mobility data elements included on other functional assessment instruments used for patients who are critically ill.

Data element Construct Validity – Observed Discharge Mobility Scores by Discharge Destination (unit of analysis is patient stays): We tested the validity of the mobility data by examining discharge function scores and whether patients were discharged to a community destination. LTCH patients who have higher abilities should be more likely to be discharged to their home or another community-based setting compared to patients discharged to another institutional post-acute care setting (e.g., skilled nursing facility, inpatient rehabilitation facility), nursing home, hospice, or an acute-care hospital. Therefore, we tested the construct validity of the mobility data by examining the relation between discharge functional abilities and the discharge

destination. We examined the relation between observed discharge mobility scores and being discharged to the community, after excluding incomplete stays.

Scale/Instrument Construct Validity – Observed Discharge Mobility Scores and Discharge Destination (unit of analysis is patient stays): We tested the validity of the scale/instrument scores by examining the observed discharge mobility scale scores and whether patients were discharged to a community destination. We ran a logistic regression model to examine the association between observed discharge mobility scores and the odds of a community discharge.

Scale/Instrument Construct Validity – Data Element (Item) Difficulty Ordering Using Rasch Analysis (unit of analysis is patient assessment data): Rasch analysis uses item data to determine how well items in a scale/instrument function together to measure a construct. In its base form, the Rasch model assumes that the probability of a code for a given item is a function of the patient's mobility ability and the item's difficulty (how hard the activity is to accomplish independently). The Rasch extension that accounts for multiple response options also considers the difficulty of moving from one code category to another (i.e., a threshold). The information resulting from this function is interval in nature and expressed on the log-odds scale. Also, as part of the analysis, Rasch methodology places persons and the items of interest on a "ruler" to enable evaluations of how well the items work together, how difficult each item is relative to the other items in the scale/instrument, and how items are ordered from easy to difficult. We used Rasch measurement analysis to examine the mobility items. We report LTCH analysis results using a Rasch-derived mobility ruler that was developed using data from LTCHs, skilled nursing facilities and inpatient rehabilitation facilities. Using the Rasch-derived cross-setting "ruler" allows comparability of mobility item functioning within and across settings.

The ordering of items from easy (bottom) to difficult (top) provides the analysis-established item difficulty hierarchy. This hierarchy can be evaluated against item design specifications (i.e., the intended construction of the items to be easy or difficult) and against expert clinical opinions as an indication of construct validity. If items are positioned into unexpected locations on the hierarchy, then the content of the items should be evaluated further and potentially modified.

Data Element (Item) and Scale/Instrument Validity - Fit Assessment Analysis (unit of analysis is patient assessment data): Rasch analysis produces fit statistics that reflect whether unexpected responses are being coded for items within the scale/instrument. The Rasch model expects the difficult items to be harder (that is, have greater need for assistance) for all patients. In a similar way, patients with higher functional abilities are generally expected to need less assistance on all items. Items that don't seem to function this way could show misfit, reflecting unexpected responses. There are two categories of fit, one designed more for outliers (outfit) and one designed for response unexpectedness near the item's difficulty (infit). In general, a cut-off appropriate for statistically determining item misfit is infit and outfit mean square values are above 1.4 when looking at multiple-point response scales. Items with fit values above 1.4 are unproductive for measurement but are not unusually "noisy" or degrade measurement. Mean square values greater than 2.0 may potentially degrade measurement (Wright and Linacre, 1994). Misfit seen near the item difficulty, or large values of infit, are concerning because they indicate noise (unexpected responses) where the item should be the most productive for measurement.

Data Element (Item) and Scale/Instrument Validity - Response Option Assessment Using Rasch Analysis (unit of analysis is patient assessment data): Rasch analysis output reports the number and percent of patients by score level (06 - Independent to 01 - Dependent) for each item and the average mobility ability (i.e., scale-level ability) of those patients. This allows us to examine if the 6-point rating scale is operating as intended for the mobility items. In general, we expect that patients who have lower ability overall would have lower ability levels (i.e., lower scores) for each item. Therefore, the average mobility ability calibration (scale-level ability measure reported in logits) associated with the more dependent scores would be lower than those associated with the more independent scores.

Citation:

Wright BD, Linacre JM (1994) Reasonable mean-square fit values. Rasch Measurement Transactions. 8:3 p.370. http://www.rasch.org/rmt/rmt83b.htm

2b1.3. What were the statistical results from validity testing? (*e.g., correlation; t-test*)

Content Validity: Similarity of Data Elements Across Other Mobility Assessment Instruments.

Many functional status scales have been developed for critically ill patients for various research and clinical uses. To address content validity, we provide a table that list activities (data elements) used to calculate the Change in Mobility performance measure and data elements included in other functional assessment scales for critically ill patients. **Table 4** shows that the Section GG mobility activities cover a wide range of mobility activities and that many of the activities included on other instruments (e.g., sit to stand, toilet transfer) are included in Section GG.

Table 4. Comparison of Selected Mobility Activities (Data Elements) for the Change in Mobility Among Patients Requiring Ventilator Support Performance Measure and Other Critical Care Functional Assessment Instruments.

Activity (Data elements)	Change in Mobility Measure: Section GG Mobility Data elements	Short Physical Performance Battery	Functional status Score – Intensive care Unit	De Morton Mobility Index	Critical Care Functional Rehabilitation Outcome Measure	Chelsea Critical care Physical Assessment Tool
Roll left and right **ICF =** Rolling over d4107	✓		✓	✓	✓ (L or R)	✓
Sit to lying **ICF =** Lying down d4100	✓					
Lying to sitting on side of bed **ICF =** Lying down d4100	✓		✓	✓	✓	✓
Sit to stand **ICF =** Standing d4104	✓	✓ 5 times	✓	✓	✓	✓
Chair/bed-to-chair transfer **ICF =** Transferring oneself while sitting d4200	✓				✓	✓
Toilet transfer **ICF =** Transferring oneself while sitting d4200	✓	✓	✓	✓	✓	✓
Walk 50 feet with two turns **ICF =** Walking and moving, other specified and unspecified d469	✓					
Walk 150 feet **ICF =** Walk short distances d4500	✓		✓			✓
Dynamic sitting			✓	✓	✓ (10 sec)	✓

Data Element Construct Validity: Observed Discharge Functional Ability and Discharge Destination (unit of analysis is patient stays). As shown in **Table 5,** patients with higher discharge scores (from 01 - Dependent to 06 – Independent) are more likely to be discharged to the community. There are two exceptions. One exception is level 01, which is slightly higher than levels 02 and 03 for bed mobility and transfer data elements,

and the second is that level 02 is slightly higher than level 03 for walk 50 feet with 2 turns. These findings may reflect that patients with incomplete stays (e.g., patients discharged to acute care) were excluded from this analysis, because discharge function data are not collected due to the urgent nature of the discharge. As expected, for each of the mobility data elements (**Table 5**), patients who were coded as 06 - Independent, a high percentage were discharged to the community (44.7% to 74.3%).

Findings and Interpretation: Mobility data elements data were positively associated with discharge destination, as expected. Specifically, we found patients who had higher observed scores at discharge were generally more likely to be discharged to a community setting, which supports the validity of the mobility data measuring functional abilities in this LTCH population.

Table 5. Observed Discharge Mobility Data Element Scores and Discharge Location (n=40,748)

	Number (Percent) Discharged to Community (Percent Bars shown)	
GG0170A3: Mobility – Roll Left and Right		
01-Dependent	1557 (10.6%)	
02-Substantial/maximal assistance	448 (8.1%)	
03-Partial/moderate assistance	567 (8.7%)	
04-Supervision or touching assistance	659 (13.5%)	
05-Setup or clean-up assistance	595 (24.5%)	
06-Independent	2955 (44.7%)	
GG0170B3: Mobility – Sit to Lying		
01-Dependent	1808 (10.3%)	
02-Substantial/maximal assistance	381 (7.5%)	
03-Partial/moderate assistance	591 (9.6%)	
04-Supervision or touching assistance	789 (16.8%)	
05-Setup or clean-up assistance	648 (28.2%)	
06-Independent	2564 (51.9%)	
GG0170C3: Mobility – Lying to Sitting on Side of Bed		
01-Dependent	1886 (10.2%)	
02-Substantial/maximal assistance	349 (7.2%)	
03-Partial/moderate assistance	577 (9.8%)	
04-Supervision or touching assistance	836 (17.9%)	
05-Setup or clean-up assistance	648 (28.7%)	
06-Independent	2485 (54.4%)	
GG0170D3: Mobility – Sit to Stand		
01-Dependent	2244 (10.0%)	
02-Substantial/maximal assistance	269 (7.1%)	
03-Partial/moderate assistance	512 (10.6%)	
04-Supervision or touching assistance	1115 (23.7%)	
05-Setup or clean-up assistance	729 (36.9%)	
06-Independent	1912 (66.3%)	

	Number (Percent) Discharged to Community (Percent Bars shown)	
GG0170E3: Mobility – Chair/Bed-to-Chair Transfer		
01-Dependent	2242 (9.6%)	
02-Substantial/maximal assistance	273 (7.6%)	
03-Partial/moderate assistance	528 (11.4%)	
04-Supervision or touching assistance	1230 (26.1%)	
05-Setup or clean-up assistance	747 (38.5%)	
06-Independent	1761 (69.0%)	
GG0170F3: Mobility – Toilet Transfer		
01-Dependent	2463 (9.7%)	
02-Substantial/maximal assistance	207 (7.8%)	
03-Partial/moderate assistance	479 (12.4%)	
04-Supervision or touching assistance	1173 (27.1%)	
05-Setup or clean-up assistance	747 (39.1%)	
06-Independent	1712 (69.3%)	
GG0170J3: Mobility – Walk 50 Feet with Two Turns		
01-Dependent	3309 (10.4%)	
02-Substantial/maximal assistance	74 (16.8%)	
03-Partial/moderate assistance	232 (14.6%)	
04-Supervision or touching assistance	1291 (33.3%)	
05-Setup or clean-up assistance	704 (50.3%)	
06-Independent	1171 (73.9%)	
GG0170K3: Mobility – Walk 150 Feet		
01-Dependent	3705 (11.0%)	
02-Substantial/maximal assistance	65 (16.5%)	
03-Partial/moderate assistance	198 (18.3%)	
04-Supervision or touching assistance	1133 (37.9%)	
05-Setup or clean-up assistance	647 (53.7%)	
06-Independent	1033 (74.3%)	

Notes: Values reported as frequency (percent); Incomplete stays are excluded.
Activity not attempted codes are not reported.

Source: RTI analysis of LTCH CARE Data Set, July 2016 – March 2018. (Program reference: 2632_validity).

Scale/Instrument Construct Validity: Observed Discharge Functional Ability and Discharge Destination (unit of analysis is patient stays): Table 6 displays the single variable logistic regression results with observed discharge mobility scale scores as the independent variable and a dichotomous dependent variable indicating whether the LTCH patient was discharged to the community or not. The mobility scale score is the sum of the 8 mobility data element scores after recoding; the discharge scale scores can range from 8 to 48. The results show that, on average, a one-unit increase in discharge mobility is associated with a 13.4 percent increase in the odds of being discharged to the community (OR = 1.134; p-value < 0.001).

Findings and Interpretation: Mobility scale/instrument scores were positively associated with discharge destination, as expected. Specifically, we found patients who had higher observed scores at discharge were more likely to be discharged to a community setting, which supports the validity of the scale/instrument data measuring functional abilities in the LTCH population.

Table 6. Coefficient and Odds Ratio for Discharge to Community Model (n=44,052)

Independent Variable	Value	95% Confidence Interval
Discharge Mobility Item Score		
Coefficient	0.126	
Odds Ratio	1.134	1.129 – 1.138

Note: Observed Discharge Mobility score range = 8 – 48; Incomplete stays were excluded.
Source: RTI analysis of LTCH CARE Data Set, July 2016 – March 2018. (Program reference: 2632_validity).

Scale/Instrument Construct Validity: Data Element (Item) Difficulty Ordering Using Rasch Analysis (unit of analysis is patient assessment data): We used Rasch analysis to determine how well the mobility items work together to measure the construct of mobility. Rasch analysis creates a mobility ruler using log odd units (i.e., logits) centered at the value 0. A "logit" (a contraction of "Log-Odds Unit") is a linear scale. We report LTCH analysis results using a Rasch-derived mobility ruler that was developed using data from LTCHs, inpatient rehabilitation facilities, and skilled nursing facilities. The analysis of the Section GG mobility data show that the placement of each mobility item on the cross-setting mobility "ruler" make sense clinically and are consistent with previous analyses of other functional assessment scale/instruments. That is, the order of items from easy to difficult (item hierarchy), is consistent with task difficulties. The order of the items by difficulty level, with the hardest activity listed first, is as follows:

Walk 150 Feet (most difficult activity)

Walk 50 Feet with Two Turns

Toilet Transfer

Chair/Bed Transfer

Sit to Stand

Lying to Sitting

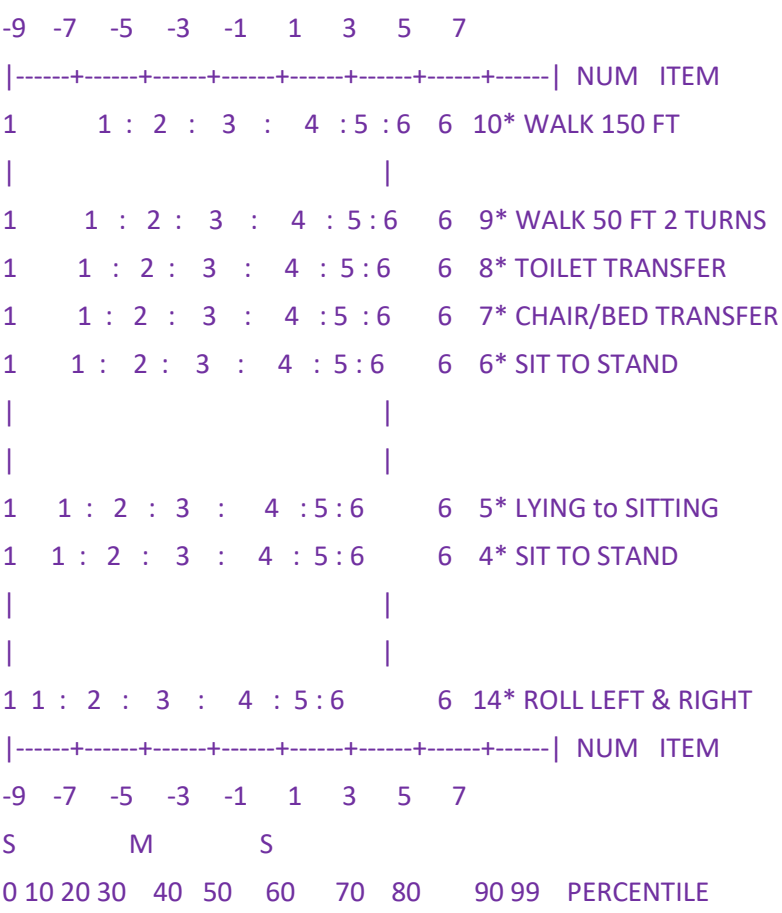
Sit to Lying

Roll Left & Right (easiest activity)

Figure 1 reports the item hierarchy, the patient distribution and the rating scale scores in one graphic. In addition, **Figure 1** is presented on the Rasch-derived mobility ruler, expressed in logits and centered at a value of 0, as described previously. It shows the overall expected score placement on the mobility "ruler" for each item. The ruler values, ranging from -9 to +7 logits, are shown on the top and bottom vertical lines. The difficulty order (item hierarchy), from easy (bottom) to difficult (top), is shown on the right side of the graphic. For each item presented on the right, the overall expected placement of the score options (from "1" for "dependent" to "6" for "independent") are shown along the ruler. Each item is presented on a row and the scores begin with the most dependent (represented by the "1") on the far-left graphic boundary and the most independent (represented by "6") on far-right graphic boundary. Finally, the threshold between two score options is represented by a colon (:) and is where a patient has an equal chance of being in either the higher or lower category. Use of the "ruler" allows visualization of the scores for each mobility item in relation to the scores of other mobility items. The letters at the bottom of **Figure 1** describe the distribution of people along the ruler, where "M" is the average of the sample and "S" and "T" are one and two times the standard deviation around that average, respectively. The percentile values represent the distribution of patients along the "ruler."

Findings and Interpretation: The item hierarchy listing and **Figure 1** illustrate that the mobility items fall along the cross-setting "ruler" as expected and are consistent with clinical findings from applications in the field and other functional assessment instruments.

Figure 1. Mobility LTCH Items – Anchored on the Cross-Setting Mobility Ruler



Scale/Instrument Validity - Fit Assessment Using Rasch Analysis (unit of analysis is patient assessment data): Ideal measurement construction would mean data fit the Rasch model exactly. In reality, empirical data will differ from the model. Rasch fit statistics describe how well the observed data (e.g. patient's scores on the mobility items) fit the model, and characterize the magnitude that unexpected scores (i.e., unmodelled noise) are found in the data. Fit statistics have an expected value of 1.0 and can range from 0 to infinity. Values lower than 1.0 indicate overfit (over prediction) of the Rasch model and values greater than 1.0 indicate underfit of the model (e.g., noise). There are two categories of fit. Outfit is designed more for outliers (when a patient's unexpected code is for an item that is relatively easy or hard for that patient); Infit is designed for unexpected codes near the item's difficulty (when a patient's code is for an item is near that person's ability). Values greater than 2.0 may potentially degrade measurement (Wright and Linacre, 1994). Overall, the mobility items are coded as expected. **Table 7** reports fit statistics for the mobility items and shows that there is one item, Walk 150 Feet with an outfit mean square above 2.00. This misfit may be due to the high acuity of the LTCH patient population.

Table 7. Fit Statistics for the Mobility Data elements (n = 323,230)

	LTCH – Anchored (Cross-Setting Ruler)	
Data element	**Infit mean square**	**Outfit mean square**
GG0170A: Roll Left & Right	1.28	1.39
GG0170B: Sit to Lying	0.75	0.77
GG0170C: Lying to Sit	0.66	0.67
GG0170D: Sit to Stand	0.73	0.75
GG0170E: Chair/Bed-to-Chair Transfer	0.76	0.79
GG0170F: Toilet Transfer	0.79	0.84
GG0170J: Walk 50 Feet with Two Turns	1.42	1.56
GG0170K: Walk 150 Feet	1.98	2.28

Data Element (Item) and Scale/Instrument Validity - Response Option Assessment Based on Rasch Analysis (unit of analysis is patient assessments): Rasch analysis provides information on how many patients are coded in each score category (i.e., independent to dependent) for each item and the average ability (or skill level) of those individuals on the construct of interest. Evaluations of patient ability by score category indicate that rating scale use is as expected, with patients with higher data element scores are, on average, higher ability patients. For our data, we anticipate that for each item, patients with higher scores (01 to 06) should have higher Rasch logit mobility values (Rasch mobility logit values range from -9 to +7). Likewise, it is expected that lower ability persons would generally be observed in the more dependent categories (e.g., substantial assistance). Therefore, the average ability (or skill level) estimate associated with the more dependent scores would be lower than ability estimates associated with the more independent scores. We combined admission and discharge data for each data element in order to ensure a range of patient ability is represented in the analyses.

As shown in **Table 8**, for each data element, patients who have higher scores have higher overall mobility ability, as expected. This is observed for each data element and each score level.

Table 8. Distribution of Combined Admission and Discharge Scores and Average Ability Estimate by Response Code (n = 321,392)

Data element	Score (Response Code)* Higher Score = Higher Ability	Number of Patients	Percent of Patients by Data Element	Average Mobility Ability of Patients (- 9 to +7 Logit Scale; Higher Value = Higher Ability)
Roll Left & Right				
	01	85,931	27	-8.36
	02	40,922	13	-5.76
	03	49,646	15	-3.49
	04	40,625	13	-0.85
	05	20,852	6	1.26
	06	83,546	26	3.94
Sit to Lying				
	01	73,037	26	-8.30
	02	36,730	13	-5.06
	03	47,268	17	-2.67
	04	38,920	14	0.09
	05	19,248	7	1.97
	06	67,883	24	4.64
Lying to Sitting on Side of Bed				
	01	67,477	25	-8.25
	02	35,396	13	-4.93
	03	46,091	17	-2.50
	04	38,989	14	0.29
	05	18,521	7	2.19
	06	64,693	24	4.77
Sit to Stand				
	01	54,417	24	-7.59
	02	25,856	12	-4.06
	03	38,487	17	-1.61
	04	41,251	19	1.26
	05	16,216	7	2.97
	06	46,247	21	5.53
Chair/Bed-to-Chair Transfer				
	01	80,280	33	-7.34
	02	24,630	10	-3.90
	03	37,213	15	-1.43
	04	41,642	17	1.40
	05	16,684	7	3.09
	06	44,620	18	5.63

Data element	Score (Response Code)* Higher Score = Higher Ability	Number of Patients	Percent of Patients by Data Element	Average Mobility Ability of Patients (- 9 to +7 Logit Scale; Higher Value = Higher Ability)
Toilet Transfer				
	01	57,761	28	-7.39
	02	20,207	10	-3.80
	03	32,397	16	-1.29
	04	38,128	18	1.48
	05	16,179	8	3.12
	06	43,268	21	5.66
Walk 50 Feet with Two Turns				
	01	7,590	8	-4.17
	02	3,339	4	-2.02
	03	10,116	11	-0.52
	04	29,186	31	1.89
	05	11,397	12	3.51
	06	31,691	34	6.02
Walk 150 Feet				
	01	9,576	12	-3.23
	02	3,075	4	-1.55
	03	7,240	9	-0.04
	04	21,953	27	2.21
	05	9,823	12	3.72
	06	28,492	36	6.13

Note: Activity not attempted/did not occur codes are not included in this analysis.

*Score are defined as: 01 – Dependent; 02 – Substantial/maximal assistance; 03 - Partial/moderate assistance; 04 - Supervision or touching assistance; 05 - Setup or clean-up assistance; and 06 - Independent.

2b1.4. What is your interpretation of the results in terms of demonstrating validity? (i.*e., what do the results mean and what are the norms for the test conducted?*)

The activities (data elements) included in the Section GG mobility scale/instrument are included in many other functional assessment instruments used for critically ill patients, supporting content validity of the scale/instrument. We found that patients who had higher discharge scores for the mobility data elements were generally more likely to be discharged to the community, as expected, and the observed mobility scale/instrument scores were significantly associated with being discharged to the community.

The difficulty order of the mobility data elements makes sense clinically and are consistent with previous analyses of the mobility data and analyses of other functional assessment scales/instruments. Rasch analysis of the data showed the data elements work well together to measure the concept of mobility, with generally good infit and outfit statistics. As expected, for each data element, the average ability score of patients increases as the rating scale/instrument increases. All of these results support the validity of the mobility data elements and scale/instrument in measuring mobility functional abilities for LTCH patients Requiring Invasive Ventilator Support.

2b2. EXCLUSIONS ANALYSIS

NA ☐ no exclusions — *skip to section* 2b3

2b2.1. Describe the method of testing exclusions and what it tests (*describe the steps—do not just name a method; what was tested, e.g., whether exclusions affect overall performance scores; what statistical analysis was used*)

We examined the number and percentage of patients who were excluded from the performance measure calculation due to exclusion criteria. The exclusion criteria are applied to the data in order to maintain the validity of the calculated performance measure scores and were identified in consultation with expert panel members and in response to public comments. Some, but not all, LTCHs admit patients with traumatic spinal cord injury and traumatic brain injury; therefore, application of these exclusion criteria is important to ensure the validity of the calculated performance scores for all LTCHs, regardless of whether the LTCH offers specialized services for these types of patients. All exclusion criteria were applied prior to our developing the risk-adjustment model.

For several exclusion criteria, the rationale for the exclusion of these patients is that improvement in mobility would be limited or unpredictable. For these exclusion criteria, we report the mean, median and 25th and 75th percentiles for change in mobility scores.

For patients who have an incomplete stay (e.g., emergency discharge), it is challenging to collect accurate discharge functional status data due to the urgent nature of the discharge. Therefore, patients with incomplete stays are excluded from the performance measure calculation, and we are unable to conduct analyses due to the unavailability of data. A total of 24,918 (37.7%) patient stays were classified as incomplete stays based on the definition of an incomplete stay.

We excluded patients younger than 21 in our original measure specifications, because we had very few patients in our sample younger than 21 and there is limited literature about functional outcomes for chronically critically ill patients younger than 21. We are maintaining this exclusion criterion, because there is still limited evidence in the literature about function outcomes for this population. A total of 435 (0.7%) patient stays were excluded due this exclusion criterion.

2b2.2. What were the statistical results from testing exclusions? (*include overall number and percentage of individuals excluded, frequency distribution of exclusions across measured entities, and impact on performance measure scores*)

A total of 33,419 patient stays (50.5%) are excluded from the change in mobility for patients requiring ventilator support performance measure. As indicated above, most of these (24,918 (37.7%)) are due to incomplete stays. An analysis of differences between patient-level characteristics for those included and excluded from the performance measure (available upon request) show little variation in the two populations. The largest difference was 2.7% for the 64 and younger patient group (44.2% for the full population and 46.9% for the population with exclusion applied). As noted above, these exclusion criteria are important to apply to ensure the validity of the calculated performance scores for all LTCHs, regardless of whether the LTCHs offers specialized services for patients with these excluded medical conditions.

Table 9 shows the number and percent of patients excluded for each exclusion criteria, and the mean, median and 25th and 75th percentile for the change in mobility scores (values are reported as units of change in mobility, possible range: -40 to 40). For patients with coma, complete tetraplegia, locked-in syndrome, severe anoxic brain damage, multiple sclerosis, Huntington's disease, Parkinson's disease, and amyotrophic lateral sclerosis showed limited improvement. Patients discharged to hospice, and patients who are independent with all mobility activities on admission also had very limited or negative improvement.

Table 9. Observed Change in Mobility Score in Mobility Units by Exclusion Criteria (N=66,137)

Exclusion Criteria	n (%)	Mean	SD	25th Percentile	50th Percentile	75th Percentile
Discharged to Hospice	1,609 (2.4%)	0.5	4.3	0	0	0
Excluded Medical Condition						
Coma	4,659 (7.0%)	2.6	6.7	0	0	0
Complete Tetraplegia	1,530 (2.3%)	1.3	4.4	0	0	0
Locked-In Syndrome	230 (0.3%)	2.2	5.6	0	0	1
Severe anoxic brain damage, cerebral edema, or compression of the brain	5,144 (7.8%)	3.8	8	0	0	4
Multiple Sclerosis	431 (0.7%)	2.8	6.9	0	0	3
Huntington's Disease	39 (0.1%)	4.7	8.5	0	1	6
Parkinson's Disease	900 (1.4%)	3.5	6.7	0	0	4
Amyotrophic Lateral Sclerosis	518 (0.8%)	2.7	6.7	0	0	3
Independent with all Admission Mobility Data elements	49 (0.1%)	-12.7	17	-36	0	0

Note: N = number of patient stays; Observed Change in Mobility values are reported as units of change in mobility (possible range: -40 to 40)

Source: RTI analysis of LTCH CARE Data Set, July 2016 – March 2018. (Program reference: 2632_exclusion)

2b2.3. What is your interpretation of the results in terms of demonstrating that exclusions are needed to prevent unfair distortion of performance results? (*i.e., the value outweighs the burden of increased data collection and analysis. Note: **If patient preference is an exclusion**, the measure must be specified so that the effect on the performance score is transparent, e.g., scores with and without exclusion*)

Patients with unexpected discharges represent the largest category of excluded cases, followed by patients with coma or severe anoxic brain damage. This is expected due to the nature of the types of patients LTCHs serve. The exclusion criteria are applied to the data in order to maintain the validity of the performance score.

2b3. RISK ADJUSTMENT/STRATIFICATION FOR OUTCOME OR RESOURCE USE MEASURES

If not an intermediate or health outcome, or PRO-PM, or resource use measure, skip to section 2b4.

2b3.1. What method of controlling for differences in case mix is used?

☐ **No risk adjustment or stratification**

☒ **Statistical risk model with 22 risk factors**

☐ **Stratification by _risk categories**

☐ **Other,**

2b3.1.1 If using a statistical risk model, provide detailed risk model specifications, including the risk model method, risk factors, coefficients, equations, codes with descriptors, and definitions.

The risk adjustment model, including the intercept (constant), covariates (risk factors) with definitions and coefficients are provided as an attached excel file and in Appendix A Table A-1. We used a Generalized Linear Model regression analysis to obtain the regression intercept (constant) and regression coefficients values.

Model for individual patient's expected change in mobility score

As described in the measure calculation algorithm, the regression intercept and coefficients are used to calculate an expected change in mobility score for each patient stay using the formula below:

The risk adjustment model includes a total of 22 covariates. For each individual patient, not every covariate will apply because, for example, only one age group, one prior indoor mobility ability, and one primary medical

condition will apply. In addition, patients could have 0 or up to 7 comorbidities. Therefore, for an individual patient stay, up to 14 covariates may apply.

Expected change in mobility score =

> intercept + (age group*coefficient) + (prior functioning: indoor ambulation*coefficient) + (prior use of wheelchair/scooter*coefficient) + (prior use of mechanical lift*coefficient) + (moderate to severe communication impairment*coefficient) + (stage 3, 4, or unstageable pressure ulcer*coefficient) + (primary medical condition*coefficient) + (comorbidity*coefficient)

In the equation above, the intercept and coefficient values were constant for each patient, while risk adjustor values were specific to the patient. Patients can have multiple comorbidities.

Risk Adjusted Change in Mobility Outcome for each LTCH

To calculate the risk adjusted change score for each LTCH, we first computed three values:

i. *Mean observed change in mobility score for each LTCH*: We calculate the mean observed change score for each LTCH as the mean of the observed change in mobility scores for all included patients treated in the LTCH.

ii. *Mean expected change in mobility score for each LTCH*: As described above, we calculate each patient's expected change in mobility score using results from the generalized linear model. We then compute the mean expected change in mobility score for each LTCH by calculating the mean of the expected change score for all included patients treated in the LTCH.

iii. *National mean observed change in mobility score*: We calculate the national mean observed change in mobility score using data for all included patients and all LTCHs.

Using the above three values, the risk adjusted change in mobility outcome for each LTCH are calculated using the formula:

$$(LTCH\ mean\ observed\ change\ in\ mobility - LTCH\ mean\ expected\ change\ in\ mobility) + National\ mean\ observed\ change\ in\ mobility$$

2b3.2. If an outcome or resource use component measure is <u>not risk adjusted or stratified</u>, provide <u>rationale and analyses</u> to demonstrate that controlling for differences in patient characteristics (case mix) is not needed to achieve fair comparisons across measured entities.

Not applicable. This performance measure is risk-adjusted.

2b3.3a. Describe the conceptual/clinical <u>and</u> statistical methods and criteria used to select patient factors (clinical factors or social risk factors) used in the statistical risk model or for stratification by risk (*e.g., potential factors identified in the literature and/or expert panel; regression analysis; statistical significance of p<0.10; correlation of x or higher; patient factors should be present at the start of care*) **Also discuss any "ordering" of risk factor inclusion**; for example, are social risk factors added after all clinical factors?

This performance measure estimates the risk adjusted mean change in mobility score between admission and discharge among LTCH patients requiring invasive ventilator support at admission. Functional improvement can vary based on patients' demographic or clinical characteristics; therefore, this measure is risk adjusted. The goal of risk adjustment is to control for differences across facilities in patient characteristics at admission that might be related to the outcome of interest. This allows outcomes to be compared across facilities after differences in patient complexity (i.e., patient characteristics) have been accounted for in the analysis. The risk adjustment model for this measure controls for variation across facilities in patient demographic (e.g., age) and clinical (e.g., diagnosis) characteristics present at the time of admission that may influence mobility outcomes, to allow change in mobility outcomes to be compared across LTCHs.

Initial development of the risk adjustment model can be found on this measure's previous testing form. We are now updating the risk adjustment model for this measure using the national data collected using the LTCH CARE Data Set, including patients' primary conditions, prior functioning, and comorbidities at admission.

Risk Adjustor Selection – Conceptual Rationale and Statistical Testing

The initial selection of risk adjustors was based on a review of the literature, input from technical experts and public comments, followed by data analysis. Please see the 2014 testing form on this measure for more detailed information on the initial selection of risk adjustors for this measure. In preparation for endorsement maintenance, we updated our literature review and conducted additional analyses.

We tested the risk adjustors using a generalized linear model with generalized estimation equations (GEE) as the estimation method to account for clustering of data within each LTCH. The generalized estimation equations method accounted for potentially correlated outcomes of patients within the same LTCH, in addition to risk adjusting the change in mobility outcome using the final set of risk adjustors.

The dependent variable was the change in mobility score for each patient, calculated as the difference between the discharge mobility score and admission mobility score. The regression coefficient represents the effect of an individual covariate. For example, a coefficient value of -0.5 for a comorbidity would be interpreted to mean that, on average, patients with that comorbidity had a change in mobility score that was 0.5 mobility units less than patients without that comorbidity.

Risk adjustors were added to the model together and decisions were made to retain or drop each risk adjustor based on its sample size, regression coefficient, significance level, and clinical relevance to mobility outcomes. For example, we dropped comorbidities that no longer showed a negative association with the dependent variable. The final risk adjustor decisions were based on a combination of clinical reasoning and statistical findings.

Risk adjustors included in the final model are described below, and also presented in S. 2b. Data Dictionary, Code Table, or Value Sets.

Age groups: We included four age groups in the risk adjustment model (< 55 years, 55-64 years, 75-84 years, and ≥ 85 years). The age group 65-74 years formed the reference category. Age was not normally distributed in our sample, so it was more appropriate to use age groups in our analyses. Patients younger than 55 years and those 55-64 years old had significantly larger change in mobility scores compared with the reference category, while patients 85 years and older had significantly smaller change in mobility scores compared with the reference category. Patients 75-84 years old also had smaller change in mobility scores (coefficient = -1.6863; $p < 0.0001$) compared with the reference group.

Communication Impairment: Communication impairment includes both expression (Expression of ideas and wants) and comprehension (Understanding verbal content) abilities. While expression and comprehension abilities are assessed and reported separately, we combined them into a single rating of communication impairment for risk adjustment. The final risk adjustment model includes “moderate to severe communication impairment” as a risk adjustor, this risk adjustor being a significant negative predictor of change in mobility outcomes (coefficient = -1.9412, $p < 0.0001$). In the final risk adjustment model, “mild to no communication impairment” forms the reference category.

Prior Functioning - Indoor Ambulation: We included patient’s functional ability in indoor ambulation prior to onset of their presenting illness, injury or exacerbation, as a risk adjustor in the model. We included separate categories for patients who were dependent in indoor ambulation, and those who needed some help in indoor ambulation prior to their current illness. Patients who were previously independent in indoor ambulation formed the reference category. Regression analyses showed that patients who were previously “dependent” in indoor ambulation, and those who needed “some help” in indoor ambulation had significantly smaller change in mobility scores compared with the reference category. The coefficient for the “dependent” category (coefficient = -4.2700, $p = < 0.0001$) was larger than that for the “some help” category (coefficient = -1.9684, $p = < 0.0001$).

Prior Mobility Devices/Aids – Wheelchair/Scooter and Mechanical Lift: We included use of a wheelchair/scooter and use of a mechanical lift prior to current illness, injury, or exacerbation, as two separate

risk adjustors in the final model. Both variables had large negative coefficients (-2.0660 for prior wheelchair use, and -2.4056 for prior mechanical lift use) and were included due to their clinical importance.

Diagnoses - Primary Medical Condition: The final risk adjustment model includes four primary medical condition categories, "chronic respiratory", "acute and chronic respiratory", "chronic cardiac", and "other primary medical condition"; "acute respiratory conditions" formed the reference category.

Stage 3, 4, or Unstageable Pressure Ulcers: We included a variable for presence of one or more stage 3, 4, or unstageable pressure ulcers, with the reference category being patients who did not have a stage 3, 4, or unstageable pressure ulcer. Patients with stage 3, 4, or unstageable pressure ulcers had significantly smaller change in mobility scores (coefficient = -1.7629, p = < 0.0001) compared with the reference category.

***Diagnoses -Comorbid Conditions*:** The final risk adjustment model includes 7 medical conditions as comorbid conditions. For each patient, the comorbid condition is either present or not present; more than one comorbid condition can apply for a single patient. The comorbid conditions are: Severe and Metastatic Cancers; Dialysis and Chronic Kidney Disease, Stage 5; Diabetes Mellitus; Major Lower Limb Amputation; Stroke, Hemiplegia or Hemiparesis; Dementia; Paraplegia, Incomplete Tetraplegia, Other Spinal Cord Disorder/Injury.

2b3.3b. How was the conceptual model of how social risk impacts this outcome developed? Please check all that apply:

☐ **Published literature**

☒ **Internal data analysis**

☐ **Other (please describe)**

2b3.4a. What were the statistical results of the analyses used to select risk factors?

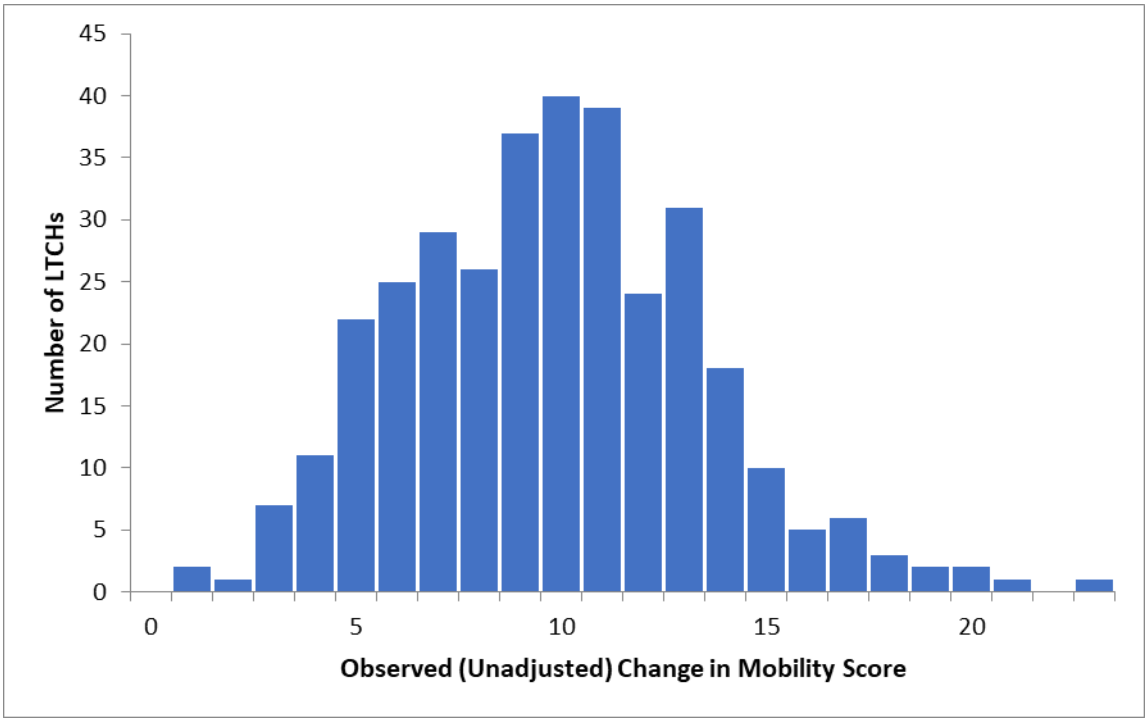
Results of the final risk adjustment model are shown in S.2b. Data Dictionary, Code Table, or Value Sets, along with regression coefficients and significance values of the final set of risk adjustors.

As described above, decisions were made to retain or drop each risk adjustor based on its sample size, regression coefficient, significance level, and clinical relevance to mobility outcomes. For example, we dropped comorbidities that no longer showed a negative association with the dependent variable, because comorbidities are expected to limit functional improvement. The final risk adjustor decisions were based on a combination of clinical reasoning and statistical findings.

The overall model was a significant predictor of change in mobility scores, with a p-value less than 0.001. The overall model R-square was 0.15, indicating that 15% of the variance in change in mobility was explained by the model. In general, regression coefficients of individual risk adjustors demonstrated that the predictive ability of risk adjustors was as clinically expected.

Distributions of the facility-level mean unadjusted and risk adjusted change in mobility scores are shown in **Figures 2 and 3 and Table 10. Figure 2** demonstrates normal distribution and good variability of the facility-level mean unadjusted change scores with some outliers. Similarly, **Figure 3** also demonstrates normal distribution and good variability of the facility-level mean risk adjusted change in mobility scores, but more concentrated around the average and fewer outliers.

Figure 2. Distribution of Facility-Level Mean Unadjusted Change in Mobility Scores (N=343)



Source: RTI analysis of LTCH CARE Data Set, July 2016 – March 2018. (Program reference: 2632_03)

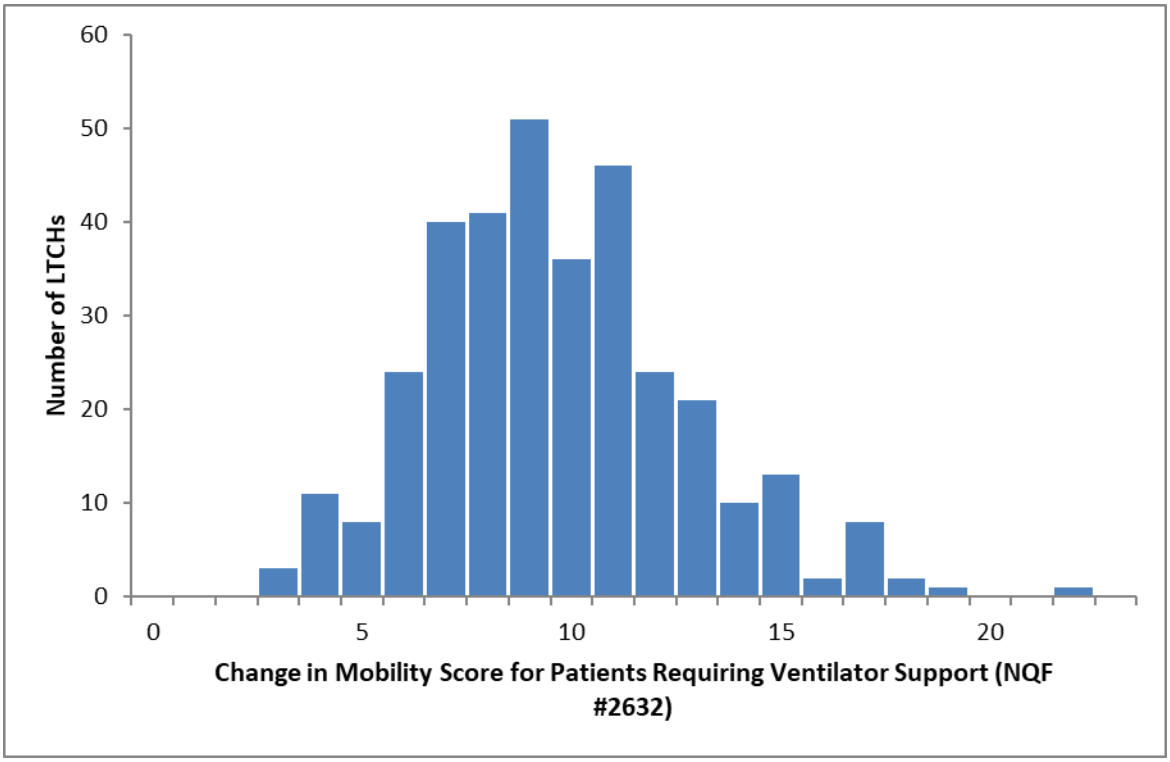
Table for Figure 2. Distribution of Facility-Level Mean Unadjusted Change in Mobility Scores (n=343)

Mean Unadjusted Change in Mobility Scores*	Number of LTCHs
1.0 to 2.0	2
2.0 to 3.0	1
3.0 to 4.0	7
4.0 to 5.0	11
5.0 to 6.0	21
6.0 to 7.0	26
7.0 to 8.0	30
8.0 to 9.0	26
9.0 to 10.0	36
10.0 to 11.0	41
11.0 to 12.0	38
12.0 to 13.0	24
13.0 to 14.0	31
14.0 to 15.0	18
15.0 to 16.0	11
16.0 to 17.0	5
17.0 to 18.0	6
18.0 to 19.0	2
19.0 to 20.0	2
20.0 to 21.0	3
21.0 to 22.0	0
22.0 to 23.0	0

Mean Unadjusted Change in Mobility Scores*	Number of LTCHs
23.0 to 24.0	1
Total	**343**

*Scores were rounded to the nearest whole number for the figure
Source: RTI analysis of LTCH CARE Data Set, July 2016 – March 2018. (Program reference: 2632_03)

Figure 3. Distribution of Facility-Level Mean Risk Adjusted Change in Mobility Scores (N=343)



Source: RTI analysis of LTCH CARE Data Set, July 2016 – March 2018. (Program reference: 2632_03)

Table for Figure 3. Distribution of Facility-Level Mean Risk-Adjusted Change in Mobility Scores (n=343)

Mean Risk Adjusted Change in Mobility Scores*	Number of LTCHs
1.0 to 2.0	0
2.0 to 3.0	0
3.0 to 4.0	3
4.0 to 5.0	11
5.0 to 6.0	8
6.0 to 7.0	24
7.0 to 8.0	40
8.0 to 9.0	41
9.0 to 10.0	50
10.0 to 11.0	38
11.0 to 12.0	44
12.0 to 13.0	25
13.0 to 14.0	21
14.0 to 15.0	11
15.0 to 16.0	13
16.0 to 17.0	2

Mean Risk Adjusted Change in Mobility Scores*	Number of LTCHs
17.0 to 18.0	7
18.0 to 19.0	3
19.0 to 20.0	0
20.0 to 21.0	0
21.0 to 22.0	0
22.0 to 23.0	1
23.0 to 24.0	0
Total	**343**

*Scores were rounded to the nearest whole number for the figure

Source: RTI analysis of LTCH CARE Data Set, July 2016 – March 2018. (Program reference: 2632_03)

Table 10 shows that the standard deviation and standard error of the mean risk adjusted change scores are similar than those of the unadjusted change scores. The mean risk adjusted change scores have a range of 2.9 to 21.9, and an interquartile range of 4.0. In contrast, the mean unadjusted change scores have a wider range of 0.2 to 22.3, and a wider interquartile range of 5.1. Skewness values of the facility-level mean unadjusted change scores are similar to those of the mean risk adjusted change scores (**Table 10**), and kurtosis values of the facility-level mean unadjusted change scores are smaller than those of the mean risk adjusted change score, indicating that the unadjusted scores deviate from a normal distribution to a smaller extent than the risk adjusted scores (computed performance measure scores).

Table 10. Distribution of Facility-Level Mean Unadjusted and Risk Adjusted Change in Mobility Scores

Change in Mobility Score	N	Mean (SD)	SE	Min	10th Pctl	25th Pctl	Median	75th Pctl	90th Pctl	Max	Skewness	Kurtosis
Unadjusted (Observed)	343	9.3 (3.6)	0.2	0.6	4.7	6.6	9.2	11.7	13.6	22.3	0.3	0.1
Risk Adjusted	343	9.2 (3.1)	0.2	2.9	5.5	7.0	8.9	11.0	13.3	21.9	0.5	0.5

N = Number; SD = Standard deviation; SE = standard error; Min = Minimum; Pctl = Percentile; Max = Maximum; Providers with < 20 stays during the 21-month testing period are excluded.

Source: RTI analysis of LTCH CARE Data Set, July 2016 – March 2018. (Program reference: 2632_03)

2b3.4b. Describe the analyses and interpretation resulting in the decision to select social risk factors *(e.g. prevalence of the factor across measured entities, empirical association with the outcome, contribution of unique variation in the outcome, assessment of between-unit effects and within-unit effects.)* **Also describe the impact of adjusting for social risk (or not) on providers at high or low extremes of risk.**

We examined whether 4 social risk factors affected performance measure scores. The social risk factors are: 1) payer source (patient-level variable); 2) marriage status (patient-level variable); 3) race (patient-level variable); and 4) ethnicity (patient-level variable).

Payer source, marriage status, race, and ethnicity data were derived from the LTCH CARE Data Set.

We conducted the following analyses to examine the effect of the 4 social risk factors:

- We calculated the percentage of stays for each social risk factor subgroup;
- We calculated the change in mobility score for each social risk factor subgroup;
- We added indicators for each social risk factor to our risk adjustment model and estimated the coefficients of these risk data element in the model; and
- We calculated the difference in provider scores with and without social risk factor adjustment.

Table 11 shows the distribution of the social risk factors from July 2016 through March 2018 and the mean change in mobility score by social risk factor subgroup.

The unadjusted mean change in mobility score varied by race/ethnicity and payer source. Black and Asian patients had lower unadjusted change in mobility scores than White patients. Medicaid patients, who were likely dual-eligible, had lower unadjusted change scores than both patients with Medicare and private insurance. Patients not currently married also had lower unadjusted changes scores than patients who are married. Mean change in mobility scores were similar when examining ethnicity.

Table 11. Distribution of Social Risk Factors and Mean Change in Mobility Score for LTCH Patients (N = 32,718)

Social Risk Factor	n	%	Observed Change in Mobility (unadjusted)
Race			
White	21,720	66.4	9.5
Black or African American	6,183	18.9	6.8
Asian	477	1.5	6.2
American Indian/Alaskan Native	210	0.6	12.7
Native Hawaiian/Pacific Islander	66	0.2	10.3
None of the Above	4,062	12.4	7.9
Ethnicity			
Hispanic or Latino	1,570	4.8	8.4
Non-Hispanic or Latino	31,148	95.2	8.8
Marriage Status			
Married	12,985	39.7	9.2
Not Currently Married	19,733	60.3	8.5
Payer Source			
Medicare Fee-for-Service	5,113	15.6	8.6
Medicare Managed Care	3,694	11.3	8.3
Medicaid Fee-for-Service	7,330	22.4	7.9
Medicaid Managed Care	2,809	8.6	8.2
Private Managed Care	2,812	8.6	9.9
Other**	10,789	33.0	9.4
Unknown	171	0.5	4.7

* Includes never married, widowed, separated, divorced, and not assessed/no information.

** Includes Private insurance/Medigap, Workers' compensation, title programs, other government sources, self-pay, or no payor source.

Source: RTI analysis of LTCH CARE Data Set, July 2016 – March 2018. (Program reference: 2632_rfa)

Table 12 shows the social risk factor coefficient estimates in our Generalized Linear regression model. Compared to patients who were White, Black and Asian patients had lower change in mobility scores though this was not significant for Asian patients. American Indian/Alaskan Native and Hispanic patients had on average higher mobility changes, while patients not currently married had slightly lower changes, on average. Compared to patients with private insurance, patients on Medicaid had on average slightly lower change in mobility scores.

Table 12. Effect of Social Risk Factors in the LTCH Change in Mobility Regression Model (N = 32,718)

Social Risk Factor	Coefficient Estimate	SE	p-value
Race			
White (reference)	--	--	--
Black or African American	-1.4954	0.1477	< 0.0001
Asian	-0.6271	0.4628	0.1754
American Indian/Alaskan Native	2.6319	0.6875	< 0.0001
Native Hawaiian/Pacific Islander	1.2048	1.2207	0.3492
None of the Above	-1.2166	0.2006	< 0.0001
Ethnicity			
Hispanic or Latino	1.3729	0.3051	< 0.0001
Marriage Status			
Not Currently Married	-0.2423	0.1187	0.0412
Payer Source			
Medicare Fee-for-Service	0.2215	0.2366	0.3492
Medicare Managed Care	0.1685	0.2532	0.5057
Medicaid Fee-for-Service	-0.4112	0.2255	0.0682
Medicaid Managed Care	-0.7209	0.2687	0.0073
Private Managed Care (reference)	--	--	--
Other**	0.3465	0.2121	0.1024
Unknown	-0.7810	0.7894	0.3225

* Includes never married, widowed, separated, divorced, and not assessed/no information.

** Includes Private insurance/Medigap, Workers' compensation, title programs, other government sources, self-pay, or no payor source.

Note: SE=Standard error

Source: RTI analysis of LTCH CARE Data Set, July 2016 – March 2018. (Program reference: 2632_rfa)

Table 13 shows the distribution of the change in mobility performance measure scores with and without social risk factor adjustment. Overall, social risk factor adjustment had minimal impact on providers' performance measure scores. The mean difference between the two sets of scores was -0.1 percentage points, and the standard deviation difference was -0.4.

Table 13: Distribution of LTCH Change in Mobility Scores with and without Adjustment for Social Risk Factors (n = 411)

Change in Mobility Scores	Mean	SD	Min	25th Pct*	Median	75th Pct	Max
Not adjusting for SRF	9.2	3.1	2.9	7.0	8.9	11.0	21.9
Adjusting for SRF	9.2	3.0	2.8	7.0	8.9	10.8	21.7
Difference in Percentage points (SRF-adjusted minus non-SRF adjusted scores)**	0.0	-0.1	-0.1	0.0	0.0	-0.2	-0.2

* Pct = percentile. SRF = social risk factors.

**Calculated as SRF-adjusted score minus non-SRF adjusted score for each facility.

Providers with < 20 stays during the 21-month testing period are excluded.

Source: RTI analysis of LTCH CARE Data Set, July 2016 – March 2018. (Program reference: 2632_rfa)

Although race, marriage status, and payer source were associated with lower changes in mobility scores, we believe that further study is needed to better understand how social risk factors can influence health outcomes. In addition, the mean and median Change in Mobility Score with and without adjusting for the social risk factors are the same.

As noted in the Assistant Secretary for Planning and Evaluation's Report to Congress entitled "*Social Risk Factors Performance under Value-Based Purchasing*" (https://aspe.hhs.gov/pdf-report/report-congress-social-risk-factors-and-performance-under-medicares-value-based-purchasing-programs), adjusting performance measures for social factors may mask disparities in the quality of care provided, which could reduce the ability to identify and reduce them. In addition, when differences in quality are related to poor performance, bias, or discrimination, adjusting performance measures could excuse the delivery of worse care to beneficiaries with social risk factors.

Therefore, we do not adjust for social risk factors in our risk adjustment model for the LTCH Change in Mobility Among Patients Requiring Ventilator Support performance measure. We will continue to monitor the impact of social risk factors on providers' performance measure scores.

2b3.5. Describe the method of testing/analysis used to develop and validate the adequacy of the statistical model or stratification approach (*describe the steps—do not just name a method; what statistical analysis was used*)

Our risk adjustment model demonstrates reasonable predictive validity for LTCH change in mobility scores. Using multiple linear regression, we conducted regression diagnostics to assess model performance, examining predictive ability and outlier influence.

Provide the statistical results from testing the approach to controlling for differences in patient characteristics (case mix) below.

If stratified, skip to 2b3.9

2b3.6. Statistical Risk Model Discrimination Statistics (*e.g., c-statistic, R-squared*)**:**

Overall, the model explained 15% of variance in change in mobility.

2b3.7. Statistical Risk Model Calibration Statistics (*e.g., Hosmer-Lemeshow statistic*):

We conducted outlier influence analysis to assess for any outlying observations that may have large or extreme effects on the change in mobility outcome, with a Cook's D score of 1.0 or higher suggesting a potentially influential observation. All Cook's D scores were less than 1.0, with the maximum score being 0.0021.

2b3.8. Statistical Risk Model Calibration – Risk decile plots or calibration curves:

To assess model performance and stability across the sample, we divided our dataset into deciles of expected values and calculated the difference between the average observed change score and the average expected change score within each decile. A difference of 0 would indicate perfect agreement between average observed and expected change scores. We expect that the risk adjusted model performance will be stable among LTCHs regardless of whether they have patients with low or high change scores on average.

As seen in **Table 14**, there was a small amount of variability in the average difference between observed and expected scores across deciles, with a range of -1.5 to 0.9, supporting model stability across the range of expected change scores and across the sample.

Table 14. Average Difference Between Observed and Expected Change in Mobility Scores Across Deciles of Expected Change Scores (n = 32,718)

Deciles of Expected Change Scores	Sample Size	Average Observed Change Score	Average Expected Change Score	Average Difference (Observed – Expected)
Decile 1 (-11.5 – 3.1)	3,272	2.0	0.5	1.5
Decile 2 (3.1 – 5.3)	3,272	3.8	4.3	-0.5
Decile 3 (5.3 – 6.9)	3,276	5.4	6.2	-0.8
Decile 4 (6.9 – 8.3)	3,257	6.7	7.6	-0.9
Decile 5 (8.3 – 9.3)	3,325	7.9	8.7	-0.8
Decile 6 (9.3 – 10.3)	3,222	9.0	9.8	-0.8
Decile 7 (10.3 – 11.3)	3,276	10.8	10.7	0.1
Decile 8 (11.3 – 12.4)	3,276	12.2	11.9	0.3
Decile 9 (12.4 – 13.7)	3,483	13.7	13.0	0.7
Decile 10 (13.7 – 15.8)	3,059	15.9	14.7	1.2
Total Sample	32,718	8.7	8.7	0.0

Note: Note: N = number of patient stays; Providers with < 20 stays during the 21-month testing period are excluded.

Source: RTI analysis of LTCH CARE Data Set, July 2016 – March 2018. (Program reference: 2632_03)

2b3.9. Results of Risk Stratification Analysis:

Not applicable – no stratification

2b3.10. What is your interpretation of the results in terms of demonstrating adequacy of controlling for differences in patient characteristics (case mix)? (i.*e., what do the results mean and what are the norms for the test conducted*)

In summary, our results demonstrate reasonable predictive ability of our risk adjustment model for LTCHs.

2b3.11. Optional Additional Testing for Risk Adjustment (*not required, but would provide additional support of adequacy of risk model, e.g., testing of risk model in another data set; sensitivity analysis for missing data; other methods that were assessed*)

None

2b4. IDENTIFICATION OF STATISTICALLY SIGNIFICANT & MEANINGFUL DIFFERENCES IN PERFORMANCE

2b4.1. Describe the method for determining if statistically significant and clinically/practically meaningful differences in performance measure scores among the measured entities can be identified (*describe the steps—do not just name a method; what statistical analysis was used? Do not just repeat the information provided related to performance gap in 1b)*

For the LTCH Change in Mobility Score performance measure, we examined whether each LTCH's calculated performance measure score (i.e., the risk-adjusted change in mobility score) was worse than, better than, or no different than the national average performance of all LTCHs. For each LTCH, we calculated the 95% confidence interval for the computed performance measure score and compared this with the national mean observed change score. Facilities whose confidence interval was lower than the national mean observed change score were considered to have worse performance than the national average. Facilities whose confidence interval was higher than the national mean observed change score were considered to have better performance than the national average. Facilities whose confidence interval overlapped with the national mean observed change score were considered to be similar to national average performance.

2b4.2. What were the statistical results from testing the ability to identify statistically significant and/or clinically/practically meaningful differences in performance measure scores across measured entities? (e.g., *number and percentage of entities with scores that were statistically significantly different from mean or some benchmark, different from expected; how was meaningful difference defined*)

Table 15 shows that for the LTCH Change in Mobility Score measure, 24.5% of LTCHs had 95% confidence intervals lower than the national mean change score, indicating worse than national average performance. 23.9% of LTCHs had 95% confidence intervals higher than the national mean change score, indicating better than national average performance. Lastly, 51.6% of LTCHs had the national mean change score within the 95% confidence interval, indicating no different than then national average performance. As shown in **Figure 3** above, the LTCH calculated performance scores (i.e., the risk-adjusted change in mobility scores) are generally normally distributed.

Table 15. Comparison of Facility-Level Measure Scores with National Average Performance for LTCH Change in Mobility Score (N = 343)

Measure Name	Facility Performance Worse than National Average N (%)	Facility Performance Better than National Average N (%)	Facility Performance No Different than National Average N (%)
Change in Mobility Score	84 (24.5%)	177 (51.6%)	82 (23.9%)

Note: Providers with < 20 stays during the 21-month testing period are excluded.

Source: RTI analysis of LTCH CARE Data Set, July 2016 – March 2018. (Program reference: 2632_03)

2b4.3. What is your interpretation of the results in terms of demonstrating the ability to identify statistically significant and/or clinically/practically meaningful differences in performance across measured entities? (i.*e., what do the results mean in terms of statistical and meaningful differences?*)

These results demonstrate the ability of the measures to discriminate among facilities based on facility-level measure performance.

2b5. COMPARABILITY OF PERFORMANCE SCORES WHEN MORE THAN ONE SET OF SPECIFICATIONS

If only one set of specifications, this section can be skipped.

Note: *This item is directed to measures that are risk-adjusted (with or without social risk factors)* ***OR*** *to measures with more than one set of specifications/instructions (e.g., one set of specifications for how to identify and compute the measure from medical record abstraction and a different set of specifications for claims or eMeasures). It does not apply to measures that use more than one source of data in one set of specifications/instructions (e.g., claims data to identify the denominator and medical record abstraction for the numerator).* ***Comparability is not required when comparing performance scores with and without social risk factors in the risk adjustment model. However, if comparability is not demonstrated for measures with more than one set of specifications/instructions, the different specifications (e.g., for medical records vs. claims) should be submitted as separate measures.***

2b5.1. Describe the method of testing conducted to compare performance scores for the same entities across the different data sources/specifications (*describe the steps—do not just name a method; what statistical analysis was used*)

Not applicable

2b5.2. What were the statistical results from testing comparability of performance scores for the same entities when using different data sources/specifications? (*e.g., correlation, rank order*)

Not applicable

2b5.3. What is your interpretation of the results in terms of the differences in performance measure scores for the same entities across the different data sources/specifications? (i.*e., what do the results mean and*

what are the norms for the test conducted)
Not applicable

2b6. MISSING DATA ANALYSIS AND MINIMIZING BIAS

2b6.1. Describe the method of testing conducted to identify the extent and distribution of missing data (or nonresponse) and demonstrate that performance results are not biased due to systematic missing data (or differences between responders and nonresponders) and how the specified handling of missing data minimizes bias (*describe the steps—do not just name a method; what statistical analysis was used*)

We ran frequencies of missing data for each mobility data element at admission and discharge as well as each of the risk adjustors after applying the exclusion criteria. Missing data on the LTCH CARE Data Set is identified as a dash (-), which is coded by providers to indicate they have "No information." Dash use is expected to be a rare occurrence and coding guidance is provided through in-person and web-based trainings, training manuals, and responses to help desk inquiries.

2b6.2. What is the overall frequency of missing data, the distribution of missing data across providers, and the results from testing related to missing data? (*e.g., results of sensitivity analysis of the effect of various rules for missing data/nonresponse; if no empirical sensitivity analysis, identify the approaches for handling missing data that were considered and pros and cons of each*)

The frequencies of missing data for each mobility data element using data from the LTCH CARE Data Set are reported in **Table 16** at admission and discharge. Across all mobility data elements, at admission and discharge, the number of cases in which the mobility item data are missing is very low at 0.1 – 0.2%.

Table 16. Mobility Data Elements: Missing Data (N=41,219)

	Admission: Not Assessed (-)	Discharge: Not Assessed (-)
Mobility Data elements		
GG0170A: Roll Left & Right	48 (0.1%)	39 (0.1%)
GG0170B: Sit to Lying	60 (0.1%)	39 (0.1%)
GG0170C: Lying to Sit	60 (0.1%)	45 (0.1%)
GG0170D: Sit to Stand	85 (0.2%)	71 (0.2%)
GG0170E: Chair/Bed-to-Chair Transfer	60 (0.1%)	46 (0.1%)
GG0170F: Toilet Transfer	93 (0.2%)	74 (0.2%)
GG0170J: Walk 50 Feet with Two Turns	39 (0.1%)	76 (0.2%)
GG0170K: Walk 150 Feet	40 (0.1%)	90 (0.2%)
Total	485 (0.1%)	480 (0.1%)

Source: RTI analysis of LTCH CARE Data Set, July 2016 – March 2018. (Program reference: 2632_02).

The frequencies of missing data for each of the risk-adjustors (available upon request) is also very low, ranging from no missing data for Age and some of the comorbidity items to 0.1% for Indoor Mobility (Ambulation). Though missing data is rare, it is still accounted for in the calculation of the risk adjustors. For example, when determining Prior Device Use (Mechanical Lift) from the GG0110C data element, a dash (-) on the LTCH CARE Data Set is considered to be "0" to indicate that the patient did not use a mechanical lift rather than dropping the patient from the performance measure calculation.

2b6.3. What is your interpretation of the results in terms of demonstrating that performance results are not biased due to systematic missing data (or differences between responders and nonresponders) and how the specified handling of missing data minimizes bias**?** (i.*e., what do the results mean in terms of supporting the selected approach for missing data and what are the norms for the test conducted; if no empirical analysis, provide rationale for the selected approach for missing data*)

There is a very small number of cases with missing data, we believe this very small percentage is unlikely to cause significant bias.

Appendix A

Table A-1. Intercept and Risk-adjustor Definitions and Covariate Values for the Change in Mobility Among Patients Requiring Ventilator Support Measure, (NQF #2632)

Risk Adjustor	Category	LTCH CARE Data Set Coding and Recoding	Intercept and Coefficients for NQF #2632 All values have 4 decimal places
Model Intercept	--	--	12.6294
Age Group	<55 years	Truncate(A0220 – A0900) = age; If age < 55 years = 1; else = 0	2.9821
Age Group	55–64 years	Truncate(A0220 – A0900) = age; If age 55-64 years = 1; else = 0	2.1077
Age Group	65–74 years (reference category)	Truncate(A0220 – A0900) = age; If age 65-74 years = 1; else = 0	--
Age Group	75–84 years	Truncate(A0220 – A0900) = age; If age 75-84 years = 1; else = 0	-1.6863
Age Group	85+ years	Truncate(A0220 – A0900) = age; If age >= 85 years = 1; else = 0	-3.3091
Communication Impairment	Moderate to Severe	= [1] (Yes) if BB0700 (Expression of ideas and wants) = [1, 2] or [1] (Yes) if BB0800 (Understanding verbal content) = [1, 2] Else = [0] (No)	-1.9412
Prior functioning: indoor ambulation	Dependent	= [1] (Yes) if GG0100B = [1] (Dependent) Else = [0] (No)	-4.2700
Prior functioning: indoor ambulation	Some help	= [1] (Yes) if GG0100B = [2] (Needed some help) Else = [0] (No)	-1.9684
Prior Device Use	Manual Wheelchair or Motorized and/or Scooter	= [1] (Yes) if GG0110A (Manual wheelchair) = [1] or GG0110B (Motorized wheelchair or scooter) = [1] Else = [0] (No)	-2.0660
Prior Device Use	Mechanical Lift	= [1] (Yes) if GG0110C (Mechanical lift) = [1] Else = [0] (No)	-2.4056
Primary Medical Condition Category	Chronic respiratory condition	= [1] (Yes) if I0050 = [2] Else = [0] (No)	-2.2277

Risk Adjustor	Category	LTCH CARE Data Set Coding and Recoding	Intercept and Coefficients for NQF #2632 All values have 4 decimal places
Primary Medical Condition Category	Acute onset and chronic respiratory conditions	= [1] (Yes) if I0050 = [3] Else = [0] (No)	-0.5331
Primary Medical Condition Category	Chronic cardiac condition	= [1] (Yes) if I0050 = [4] Else = [0] (No)	-1.2701
Primary Medical Condition Category	Other medical condition	= [1] (Yes) if I0050 = [5] Else = [0] (No)	-0.8384
Stage 3, 4, or unstageable pressure ulcer/injury	Presence	= [1] (Yes) if ([M0300C1 (Number of stage 3 pressure ulcers) > 0] or [M0300D1 (Number of stage 4 pressure ulcers) > 0] or [M0300E1 (Number of unstageable pressure ulcers due to non-removable dressing/device) > 0] or [M0300F1 (Number of unstageable pressure ulcers due to coverage of wound bed by slough and/or eschar) > 0] or [M0300G1 ((Number of unstageable pressure ulcers with suspected deep tissue injury in evolution) > 0]) Else = [0] (No)	-1.7629
Comorbidities	Severe and Metastatic Cancers	= [1] (Yes) if I0103 = [1] or I0104 = [1] Else = [0] (No)	-0.1293
Comorbidities	Dialysis and Chronic Kidney Disease, Stage 5	= [1] (Yes) if O0100J = [1] or I1501 = [1] Else = [0] (No)	-0.6848
Comorbidities	Diabetes Mellitus (DM)	= [1] (Yes) if I2900 = [1] Else = [0] (No)	-0.5808
Comorbidities	Major Lower Limb Amputation	= [1] (Yes) if I4100 = [1] Else = [0] (No)	-1.7373
Comorbidities	Stroke, Hemiplegia or Hemiparesis	= [1] (Yes) if I4501 = [1] or I4900 = [1] Else = [0] (No)	-3.5778
Comorbidities	Dementia	= [1] (Yes) if I4801 = [1] Else = [0] (No)	-1.3576
Comorbidities	Paraplegia, Incomplete Tetraplegia, Other Spinal Cord Disorder/Injury	= [1] (Yes) if I5000 = [1] or I5102 = [1] or I5110 = [1] Else = [0] (No)	-5.3440

Source: RTI analysis of LTCH CARE Data Set, July 2016 – March 2018. (Program reference: 2632_03).

Appendix B:
Critical Data Element Reliability and Validity Testing

B.1 Overview of Reliability and Validity Testing

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

B.2 Traditional Inter-rater Reliability Study

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

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

- Eating: 401
- Oral hygiene: 414
- Toilet hygiene: 416
- Upper body dressing: 420
- Lower body dressing: 413
- Lying to sitting on the side of the bed: 412
- Sitting to standing: 387
- Chair/bed to chair transfer: 392
- Toilet transfer: 361
- Walk 150 feet: 68
- Walk once standing: 52
- Wheel in room: 46

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

For categorical items, kappa statistics (kappa) indicate the level of agreement between raters using ordinal data, taking into account the role of chance agreement. The ranges commonly used to judge reliability based

on kappa are as follows: ≤ 0 = poor; 0.01–0.20 = slight; 0.21–0.40 = fair; 0.41–0.60 = moderate; 0.61–0.80 = substantial; and 0.81–1.00 = almost perfect.

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

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

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

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

B.3 Videotaped Standardized Patients Reliability Study

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

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

---

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

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

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

Two main analytic approaches were used for assessing the video reliability of the CARE items, adhering closely to the methods used by Fricke et al.3 in their video reliability study of the FIM®4 instrument. First, percent agreement with the mode response was calculated for each CARE item included in at least one of the nine videos. Unlike the approach used by Fricke et al., RTI did not consider agreement at one response level above and below the mode, and instead used a stricter approach looking at direct modal agreement only. In the second approach, percent agreement with the internal clinical team's consensus response was also calculated. This second measure not only gives an indication of item reliability, but also reflects training consistency for the providers.

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

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

B.4 Scale-level Reliability Results: Internal Consistency

In addition to item-level reliability testing, we examined internal consistency, which provides a general assessment of how well the items interrelate within a domain or subscale. Internal consistency is assessed using the Cronbach's alpha coefficient, which is the average correlation of all possible half-scale divisions. Cronbach's alpha is a statistic frequently assessed when instrument or scale psychometrics are published. The Cronbach's alpha reliability estimate ranges from zero to one, with an estimate of zero indicating that there is no consistency of measurement among the items, and one indicating perfect consistency. Many cutoff criteria exist to determine whether or not a scale shows good consistency or whether the items "hang together" well. General consensus is that Cronbach's alpha should be at least 0.70 for an adequate scale for group-level decisions, and alphas closer to 1 indicate a good scale.5

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

---

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

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

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

one provider type appears to have reliability estimates higher or lower than the rest, indicating similarity of CARE usage with respect to internal consistency.
The following table shows the findings from the Cronbach's alpha internal consistency evaluation mentioned above.

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

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

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

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

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

Rasch analysis has been used to examine the FIM® instrument,[6,7,8,9] the Minimum Data Set (MDS),[10] and the Outcome and Assessment Information Set (OASIS).[11] Rasch analysis has also been used to examine the extent to which existing functional assessment instruments (e.g., the FIM® instrument, MDS 2.0) capture the same construct.[12]

---

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

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

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

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

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

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

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

Rasch measurement is based on a probabilistic model that describes the association between a person's underlying ability level and probability of a particular item response, and summarizes a patient's position along a "ruler" that represents a latent trait or concept (e.g., self-care or mobility).13 In essence, the Rasch analysis creates a ruler based on the domain measured (e.g., mobility) that can be used to assess the abilities of the patients. The analysis also provides information on the hierarchy of item difficulty (from easy to hard) that can be used to evaluate the construct validity of a set of items. In addition, the Rasch analysis provides information about the level of challenge associated with each item rating scale ("dependent" through "independent"). For example, an item with a low difficulty estimate (e.g., eating) would be more likely to be completed with little or no help by patient's items that are more challenging (e.g., 12 steps), where most patients would find completing this activity challenging. Finally, the Rasch analysis can provide information on items that do not fit into the single theorized concept through "item misfit" statistics, which may indicate that the item needs further evaluation before it is included on future administrations of the subscale. The infit mean square is an indicator of the degree to which patient responses are similar to what would be expected (i.e., predicted) by the measurement model. The acceptable range is generally 0.6 to 1.4. If the item values are above this range, it reflects that person response patterns are erratic, generally suggesting that the item is not measuring the same construct as other items. Infit mean squares above 1.4 are considered to be unacceptably unexpected14 and indicate that the item most likely does not reflect the same construct as the other items included in the scale; for example, a need for assistance with self-care.

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

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

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

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

---

[13] Wright BD, Stone MH. Best Test Design. Rasch Measurement. 1979.

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

3. Feasibility

Extent to which the specifications including measure logic, require data that are readily available or could be captured without undue burden and can be implemented for performance measurement.

3a. Byproduct of Care Processes

For clinical measures, the required data elements are routinely generated and used during care delivery (e.g., blood pressure, lab test, diagnosis, medication order).

3a.1. Data Elements Generated as Byproduct of Care Processes.

Generated or collected by and used by healthcare personnel during the provision of care (e.g., blood pressure, lab value, diagnosis, depression score), Abstracted from a record by someone other than person obtaining original information (e.g., chart abstraction for quality measure or registry)

If other:

3b. Electronic Sources

The required data elements are available in electronic health records or other electronic sources. If the required data are not in electronic health records or existing electronic sources, a credible, near-term path to electronic collection is specified.

3b.1. To what extent are the specified data elements available electronically in defined fields (*i.e., data elements that are needed to compute the performance measure score are in defined, computer-readable fields*) Update this field for **maintenance of endorsement**.

ALL data elements are in defined fields in electronic clinical data (e.g., clinical registry, nursing home MDS, home health OASIS)

3b.2. If ALL the data elements needed to compute the performance measure score are not from electronic sources, specify a credible, near-term path to electronic capture, OR provide a rationale for using other than electronic sources. For **maintenance of endorsement**, if this measure is not an eMeasure (eCQM), please describe any efforts to develop an eMeasure (eCQM).

Not applicable. This quality measure's data elements are collected solely from electronic sources.

3b.3. If this is an eMeasure, provide a summary of the feasibility assessment in an attached file or make available at a measure-specific URL. Please also complete and attach the NQF Feasibility Score Card.

Attachment:

3c. Data Collection Strategy

Demonstration that the data collection strategy (e.g., source, timing, frequency, sampling, patient confidentiality, costs associated with fees/licensing of proprietary measures) can be implemented (e.g., already in operational use, or testing demonstrates that it is ready to put into operational use). For eMeasures, a feasibility assessment addresses the data elements and measure logic and demonstrates the eMeasure can be implemented or feasibility concerns can be adequately addressed.

3c.1. Required for maintenance of endorsement. Describe difficulties (as a result of testing and/or operational use of the measure) regarding data collection, availability of data, missing data, timing and frequency of data collection, sampling, patient confidentiality, time and cost of data collection, other feasibility/implementation issues.

IF instrument-based, consider implications for both individuals providing data (patients, service recipients, respondents) and those whose performance is being measured.

The NQF feasibility criterion requires measure developers to: 1) demonstrate that the data collection strategy can be implemented and 2) describe any difficulties regarding data collection.

Data Collection:

Data for this quality measure are currently collected and submitted to the Centers for Medicare and Medicaid Services using the Long-Term Care Hospital Continuity Assessment Record and Evaluation Data Set (LTCH CARE Data Set). These data have been collected by all LTCHs in the US since April 1, 2016 as part of the LTCH Quality Reporting Program (QRP).

The measure data are "generated" by qualified clinicians as they observe patients completing daily activities, such as moving around in bed and transferring out of bed at the time of admission and discharge. As shown in the testing form, missing data is minimal (less than 0.2% across all data elements). The LTCH CARE Data Set data are submitted to CMS via the QIES ASAP system. This data submission system is secure and encrypted with administrative, physical and technical safeguards in place.

Preventing and Addressing Potential Data Collection Challenges:

The Centers for Medicare and Medicaid Services finalized the implementation of this quality measure in August 2014 in the FY 2015 IPPS/LTCH PPS Final Rule, more than 1 year before implementation of data collection. This advance notice allowed providers, vendors and CMS to prepare for implementation. The Centers for Medicare and Medicaid has developed software that is free for LTCHs to use to submit LTCH CARE Data Set data. Also, given the Centers for Medicare and Medicaid's many years of experience with data submission, implementation occurred with minimal difficulty.

To assist providers with the collection of accurate data, the Centers for Medicare and Medicaid Services has offered multiple in-person and on-line training opportunities since November 2015. In addition, a help desk is available to answer provider questions regarding data collection, and "Q & A" documents are posted on the CMS website for provider use. Training information is available on the following website: https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/LTCH-Quality-Reporting/LTCH-Quality-Reporting-Training.html.

3c.2. Describe any fees, licensing, or other requirements to use any aspect of the measure as specified *(e.g., value/code set, risk model, programming code, algorithm)*.

There no costs associated with fees, licensing or other requirements associated with the measure data elements or risk model.

4. Usability and Use

Extent to which potential audiences (e.g., consumers, purchasers, providers, policy makers) are using or could use performance results for both accountability and performance improvement to achieve the goal of high-quality, efficient healthcare for individuals or populations.

4a. Accountability and Transparency

Performance results are used in at least one accountability application within three years after initial endorsement and are publicly reported within six years after initial endorsement (or the data on performance results are available). If not in use at the time of initial endorsement, then a credible plan for implementation within the specified timeframes is provided.

4.1. Current and Planned Use

NQF-endorsed measures are expected to be used in at least one accountability application within 3 years and publicly reported within 6 years of initial endorsement in addition to performance improvement.

Specific Plan for Use	Current Use (for current use provide URL)
	Public Reporting Measure data from calendar year 2019 (currently being collected) will be publicly reported on LTCH Compare in 2020 for the LTCH Quality Reporting Program (LTCH QRP) https://www.medicare.gov/longtermcarehospitalcompare/ Quality Improvement (external benchmarking to organizations) LTCH QRP: On confidential feedback reports and LTCH Compare, providers can view national-level performance measure scores for benchmarking quality efforts. LTCHs can also review and compare scores for local providers through LTCH Compare's web features. https://qtso.cms.gov/ Quality Improvement (Internal to the specific organization) LTCH QRP: LTCHs receive confidential feedback reports through the CMS designated data submission system, which includes the Review and Correct, Quality Measure, and Provider Preview Reports to review their data internally. https://qtso.cms.gov/

4a1.1 For each CURRENT use, checked above (update for maintenance of endorsement), provide:

- Name of program and sponsor
- Purpose
- Geographic area and number and percentage of accountable entities and patients included
- Level of measurement and setting

Name of Program and Sponsor and Purpose:
This quality measure has been implemented in the Center for Medicare and Medicaid's (CMS) Long-Term Care Hospital Quality Reported Program (LTCH QRP) and serves two purposes:
1) to share quality data with each LTCH that may be used to support quality improvement efforts; and
2) to share quality data about each LTCH with the public, which may assist consumers and family members in making decisions about where to receive LTCH care.
As part of the LTCH QRP, LTCHs have been able to view data for this quality measure in their confidential feedback reports, which may be used for quality improvement, since April 2017.
Quality measure data collected in calendar year 2018 and 2019 will be publicly reported in 2020 on CMS's LTCH Compare website at: https://www.medicare.gov/longtermcarehospitalcompare. Since 2016, CMS has publicly reported LTCH QRP quality measure data on the LTCH Compare website. This website reports quality data for each LTCH, and these data are also publicly available for download at: https://data.medicare.gov/data/long-term-care-hospital-compare.
This measure was implemented pursuant to two public laws that addressed the LTCH QRP and reporting of data submitted by providers:

1) The Patient Protection and Affordable Care Act ("ACA") of 2010 (Public Law No: 111-148)
    - Section 3004(a) of the ACA amended section 1886(m)(5)(E) of the Social Security Act (SSA) requiring the Secretary to establish quality reporting requirements for LTCH providers.
    - The ACA mandates LTCHs to submit data or be subject to a two-percent reduction in their annual payment update (APU) determination.
2) The Improving Medicare Post-Acute Care Transformation Act ("IMPACT Act") of 2014 (Public Law No: 113-185):
    - The IMPACT Act requires LTCHs to submit standardized patient assessment data on quality, resource use, and other measures.
    - The data submitted from providers are used to calculate measures that report healthcare processes and patient outcomes among LTCH providers under the QRP.

- Requires the establishment of procedures for making provider performance information available to the public.

CMS finalized in the FY 2018 IPPS/LTCH PPS final rule (82 FR 38458) that they plan to publicly report data for this performance measure on LTCH Compare in the fall of 2020. The first time the data will be publicly displayed will be for patients discharged on January 1, 2018 through December 31, 2019.
CMS provides an opportunity for LTCHs to review their own data before it is publicly reported through confidential feedback reports available in the CMS designated data submission system. Several reports are available that provide different snapshots of the measure data (described in more detail below in 4a2.1.1). As of April 2017, providers could view the observed change in mobility performance measure in their confidential Review and Correct reports. The risk-adjusted change in mobility performance measure became available in the Quality Measure reports October 2017.
Geographic Area, Accountable Entities and Patients Included:
The LTCH QRP measures are calculated for 100% of LTCH providers in the US (415 LTCHs in FY 2018). LTCHs submitted a total of 102,468 pairs (admission and discharge) of LTCH CARE Data Set assessments for patients discharged in FY 2018.
All providers receive their confidential feedback reports, which may be used for internal quality improvement efforts.
To ensure reliability of the performance measure scores, LTCHs with less than 20 patients during a reporting period would not have their data displayed publicly. Once an LTCH has more than 20 patients during the reporting period, their data would display on LTCH Compare.
Level of Measurement and Setting:
As mentioned, this quality measure has been implemented in the LTCH setting as part of the LTCH QRP. The measure score is reported at the facility-level.

4a1.2. If not currently publicly reported OR used in at least one other accountability application (e.g., payment program, certification, licensing) what are the reasons? (*e.g., Do policies or actions of the developer/steward or accountable entities restrict access to performance results or impede implementation?*)

Not applicable because public reporting is currently underway for this measure.

4a1.3. If not currently publicly reported OR used in at least one other accountability application, provide a credible plan for implementation within the expected timeframes -- any accountability application within 3 years and publicly reported within 6 years of initial endorsement. (*Credible plan includes the specific program, purpose, intended audience, and timeline for implementing the measure within the specified timeframes. A plan for accountability applications addresses mechanisms for data aggregation and reporting.*)

Not applicable because public reporting is currently underway for this measure.

4a2.1.1. Describe how performance results, data, and assistance with interpretation have been provided to those being measured or other users during development or implementation.

How many and which types of measured entities and/or others were included? If only a sample of measured entities were included, describe the full population and how the sample was selected.

For Providers:

Dissemination of performance results and assistance with interpretations of the performance data for LTCHs have been addressed in four specific ways: confidential feedback reports, provider training seminars, manuals and materials, responses to questions submitted to the LTCH QRP Help Desk: LTCHQualityQuestions@cms.hhs.gov, and LTCH Public Reporting Help Desk: LTCHPRquestions@cms.hhs.gov, and on LTCH Compare.

1) Confidential Provider Feedback Reports:

All LTCHs who submit LTCH CARE Data Set data to CMS receive three types of confidential reports with performance measure data and scores based on the data submitted. These reports support internal quality improvement efforts and include the Review and Correct, Quality Measure, and Provider Preview Reports. Details about each of these reports is provided below in 4a.2.1.2.

2) LTCH QRP Provider Training Seminars:

CMS conducted several in-person LTCH QRP provider training seminars to share information about coding the data elements used to calculate the performance measure, to share details about the measure specifications and to explain how the measure is calculated. Training sessions that focused on the confidential feedback reports were also conducted to support providers in reviewing and interpreting the data they receive in these reports. During training sessions, providers were encouraged to ask questions about coding the data elements and the change in mobility performance measure to ensure an accurate understanding of the measure. Training materials are posted on the CMS website after each training seminar is completed. To review provider training materials, see the following webpage:

https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/LTCH-Quality-Reporting/LTCH-Quality-Reporting-Training.html

3) LTCH QRP and LTCH Public Reporting Help Desk:

CMS also maintains a provider help desk for the LTCH QRP where LTCHs can submit questions about the data elements, the measure, including questions about performance data, interpretation of results, or instructions on coding (LTCHQualityQuestions@cms.hhs.gov). A help desk for questions about the data available on LTCH Compare (see below) is also available (LTCHPRquestions@cms.hhs.gov). A response is provided to address each question that is submitted.

4) LTCH Compare Website:

The performance measure data are publicly displayed on the LTCH Compare website and plain language is used to assist users in interpreting the data that are presented. The quality of care that LTCH providers deliver to patients can vary from facility to facility, and publicly displaying performance data on LTCH Compare supplies information for providers to use for improving the quality of care they provide to patients.

The LTCH QRP Measure Calculations and Reporting User's manual, which presents the measure specifications and how the measures are calculated for each measure in the LTCH QRP, is posted on the CMS website. Therefore, providers have detailed measure specifications available to them. To review the current LTCH QRP Measure Calculations and Reporting User's manual, see the following webpage:

https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/LTCH-Quality-Reporting/Downloads/LTCH-Measure-Calculations-and-Reporting-Users-Manual-V30.pdf

For Patients, Families, Caregivers and Other Stakeholders:

LTCH patients, family members, caregivers, and other stakeholders (researchers, journalists, policymakers) can view an LTCH's measure performance information on the publicly available LTCH Compare website. The LTCH Compare website is designed to help patients and caregivers make informed decisions about their health care and to compare long-term care hospitals based on important indicators of quality. Preparations to include the performance data for this measure on the LTCH Compare Website includes developing plain language to explain the measure and the results for the general public. Additionally, the LTCH Compare Website has gone through consumer testing to test functionality and usability. LTCH Compare is available in both English and Spanish.

Furthermore, the public can download the LTCH Compare datasets. The files contain general information about providers, provider level data on quality measures, and national data shown on the site. A data dictionary provides detailed information on the measures and file layouts.

Public access to the performance data on the LTCH Compare website has been widespread and increasing over time. In Quarter 4 of 2017, there were 5,686 sessions and 45% of those were returning visitors. Subsequently, the number of sessions increased by 22% a year later to 6,936 sessions in Quarter 4 of 2018 in which 47% of those were returning visitors.

4a2.1.2. Describe the process(es) involved, including when/how often results were provided, what data were provided, what educational/explanatory efforts were made, etc.

All LTCHs receive three types of confidential reports with performance measure data and scores based on the data submitted:

1) Quality Measure Reports:

The intent of this report is to enable LTCHs to track their own quality measure data at the facility- and patient-level. Data for this report is refreshed monthly and displays performance measure information at the facility- and patient-stay level for review. The facility-level report displays the measure denominator, average observed scores, average risk-adjusted score, and the national average for benchmarking the facility's performance. The patient-level report displays which patients are excluded from the measure as well as each patient's observed change in mobility score.

2) Review and Correct Reports:

The intent of this report is for LTCHs to view their data prior to the quarterly data submission deadline to ensure accuracy of the data submitted to CMS. Data for this report is refreshed weekly and displays data correction deadlines and whether the data correction period is open or closed. For most measures, only the last four quarters of data are available in this report; however, the performance data for the change in mobility score for patients requiring ventilator support displays the last eight quarters of data.

3) Provider Preview Reports:

The intent of this report is for LTCHs to preview what performance data will publicly displayed for their LTCH. The report displays facility-level performance measure data and shows risk-adjusted values and national rates as they will appear publicly on LTCH Compare. Data displayed in this report cannot be modified by the provider.

4a2.2.1. Summarize the feedback on measure performance and implementation from the measured entities and others described in 4d.1.

Describe how feedback was obtained.

In addition to the processes and information described above in 4a2.1.1 and 4a2.1.2, CMS solicited public comments about the measure change in mobility among patients requiring ventilator support performance measure via a 60-day public comment period during the fiscal year (FY) 2015 rulemaking process. CMS again solicited public comments during the FY 2018 rulemaking process before finalizing this measure for public display on LTCH Compare.

FY 2015: https://www.federalregister.gov/documents/2014/08/22/2014-18545/medicare-program-hospital-inpatient-prospective-payment-systems-for-acute-care-hospitals-and-the

FY 2018: https://www.federalregister.gov/documents/2017/08/14/2017-16434/medicare-program-hospital-inpatient-prospective-payment-systems-for-acute-care-hospitals-and-the

4a2.2.2. Summarize the feedback obtained from those being measured.

We received support for both implementation and public reporting of the change in mobility among patients requiring ventilator support performance measure for the LTCH QRP. Comments were received from various stakeholders, including providers, provider associations, researchers, government agencies, information system vendors, advocacy groups, and individuals/consumers.

In the FY 2015 rule proposal, most public commenters supported the change in mobility among patients requiring ventilator support performance measure being added to the LTCH QRP. Commenters stated that functional improvement is an important patient-centered outcome especially patients who require ventilator support. Commenters also noted that such improvements would improve quality of life and reduce the likelihood of infection, morbidity, and mortality. Additionally, commenters agreed with the MAP's recommendation to adopt functional status measures and supported this measure to address the measurement gap in ventilator care.

4a2.2.3. Summarize the feedback obtained from other users

In March 2017, the measure developer convened groups of stakeholders and experts who contributed direction and thoughtful input for LTCH QRP measure development and maintenance. Some members of this technical expert panel expressed support for this measure and the importance of risk-adjustment due to the medical complexity of the LTCH population.

The LTCH QRP TEP Summary report is available at: https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/LTCH-Quality-Reporting/Downloads/LTCH-QRP-TEP-Summary-Report-December-2017.pdf

4a2.3. Describe how the feedback described in 4a2.2.1 has been considered when developing or revising the measure specifications or implementation, including whether the measure was modified and why or why not.

Part of our measure maintenance process includes incorporating stakeholder feedback as we continue examination and refinement of performance measures. CMS and RTI International reviewed and took into consideration all public comments received in the FY 2015 final rule as well as feedback from the March 2017 technical expert panel.

There was no feedback that suggested changes to the measure specifications.

Improvement

Progress toward achieving the goal of high-quality, efficient healthcare for individuals or populations is demonstrated. If not in use for performance improvement at the time of initial endorsement, then a credible rationale describes how the performance results could be used to further the goal of high-quality, efficient healthcare for individuals or populations.

4b1. Refer to data provided in 1b but do not repeat here. Discuss any progress on improvement (trends in performance results, number and percentage of people receiving high-quality healthcare; Geographic area and number and percentage of accountable entities and patients included.)

If no improvement was demonstrated, what are the reasons? If not in use for performance improvement at the time of initial endorsement, provide a credible rationale that describes how the performance results could be used to further the goal of high-quality, efficient healthcare for individuals or populations.

The change in mobility among patients requiring ventilator support performance measure was recently implemented on April 1, 2016 and will be publicly reported for the first time in the fall of 2020 using calendar year 2018 and 2019 data. Thus, there is no extensive data to evaluate trends in performance over time since this is a 24-month measure (eight quarters of data). In Section 1b, we provide analysis showing trends for rolling four quarters as well as data by quarter and show that the measure remained stable over this period. As more data becomes available, we will examine score distribution and change in provider performance scores.

4b2. Unintended Consequences

The benefits of the performance measure in facilitating progress toward achieving high-quality, efficient healthcare for individuals or populations outweigh evidence of unintended negative consequences to individuals or populations (if such evidence exists).

4b2.1. Please explain any unexpected findings (positive or negative) during implementation of this measure including unintended impacts on patients.

No unexpected findings have been identified during implementation and testing of this measure. To date, no unintended impacts on patients have been identified.

4b2.2. Please explain any unexpected benefits from implementation of this measure.

To date, no unexpected findings have been identified.

5. Comparison to Related or Competing Measures

If a measure meets the above criteria and there are endorsed or new related measures (either the same measure focus or the same target population) or competing measures (both the same measure focus and the same target population), the measures are compared to address harmonization and/or selection of the best measure.

5. Relation to Other NQF-endorsed Measures

Are there related measures (conceptually, either same measure focus or target population) or competing measures (conceptually both the same measure focus and same target population)? If yes, list the NQF # and title of all related and/or competing measures.

Yes

5.1a. List of related or competing measures (selected from NQF-endorsed measures)

0167 : Improvement in Ambulation/locomotion

0175 : Improvement in bed transferring

0422 : Functional status change for patients with Knee impairments

0423 : Functional status change for patients with Hip impairments

0424 : Functional status change for patients with Foot and Ankle impairments

0425 : Functional status change for patients with lumbar impairments

0428 : Functional status change for patients with General orthopaedic impairments

0429 : Change in Basic Mobility as Measured by the AM-PAC:

0688 : Percent of Residents Whose Need for Help with Activities of Daily Living Has Increased (long stay)

2287 : Functional Change: Change in Motor Score

2321 : Functional Change: Change in Mobility Score

2612 : CARE: Improvement in Mobility

2634 : Inpatient Rehabilitation Facility (IRF) Functional Outcome Measure: Change in Mobility Score for Medical Rehabilitation Patients

2636 : Inpatient Rehabilitation Facility (IRF) Functional Outcome Measure: Discharge Mobility Score for Medical Rehabilitation Patients

2643 : Average change in functional status following lumbar spine fusion surgery

2653 : Average change in functional status following total knee replacement surgery

2774 : : Functional Change: Change in Mobility Score for Skilled Nursing Facilities

2775 : Functional Change: Change in Motor Score for Skilled Nursing Facilities

2776 : Functional Change: Change in Motor Score in Long Term Acute Care Facilities

2778 : Functional Change: Change in Mobility Score for Long Term Acute Care Facilities

5.1b. If related or competing measures are not NQF endorsed please indicate measure title and steward.

Change in Basic Mobility as Measured by the AM-PAC (CREcare)

5a. Harmonization of Related Measures

The measure specifications are harmonized with related measures;

OR

The differences in specifications are justified

5a.1. If this measure conceptually addresses EITHER the same measure focus OR the same target population as NQF-endorsed measure(s):

Are the measure specifications harmonized to the extent possible?

Yes

5a.2. If the measure specifications are not completely harmonized, identify the differences, rationale, and impact on interpretability and data collection burden.

Quality measure NQF #0688 applies to long-stay nursing home residents, a different population. Quality measures NQF #0167, and NQF #0175use a single function activity to indicate whether patients have made functional improvement. These measures apply to home health patients, which is a different target population. Quality measures NQF #2287 and NQF #2321 apply to inpatient rehabilitation facility patients, and quality measures NQF #2774, NQF #2775 apply to skilled nursing facility patients, which are different populations. NQF measures # 0422, #0423, #0424, #0425, #and #0428 apply to outpatients, which is a different population. Quality measures NQF #2776 and NQF #2778 apply to all LTCH patients, not just patients requiring ventilator support. Quality measures NQF #2612, NQF #2634 and NQF #2636 use some of the same data elements (Section GG) that are used for this quality measure (NQF #2632).

5b. Competing Measures

The measure is superior to competing measures (e.g., is a more valid or efficient way to measure);

OR

Multiple measures are justified.

5b.1. If this measure conceptually addresses both the same measure focus and the same target population as NQF-endorsed measure(s):

Describe why this measure is superior to competing measures (e.g., a more valid or efficient way to measure quality); OR provide a rationale for the additive value of endorsing an additional measure. (Provide analyses when possible.)

Not applicable

Appendix

A.1 Supplemental materials may be provided in an appendix. All supplemental materials (such as data collection instrument or methodology reports) should be organized in one file with a table of contents or bookmarks. If material pertains to a specific submission form number, that should be indicated. Requested information should be provided in the submission form and required attachments. There is no guarantee that supplemental materials will be reviewed.

Attachment **Attachment:** LTCH_Detailed_Function_QM_Specifications_2632_01-07-2019.docx

Contact Information

Co.1 Measure Steward (Intellectual Property Owner): Centers for Medicare & Medicaid Services

Co.2 Point of Contact: Helen, Dollar-Maples, Helen.Dollar-Maples@cms.hhs.gov, 410-786-7214-

Co.3 Measure Developer if different from Measure Steward: RTI International

Co.4 Point of Contact: Anne, Deutsch, adeutsch@rti.org

Additional Information

Ad.1 Workgroup/Expert Panel involved in measure development

Provide a list of sponsoring organizations and workgroup/panel members' names and organizations. Describe the members' role in measure development.

This quality measure was developed with significant and ongoing input by several Technical Expert Panels (TEPs). Expert panel members provided input on status quality metrics, including the performance score, the target population, and exclusion criteria. Some expert panel meetings focused on measuring functional status across post-acute care settings, and other meetings focused on functional assessment and functional outcomes for LTCH patients.

Most recently, RTI International, on behalf of the Centers for Medicare & Medicaid Services (CMS), convened a Technical Expert Panel (TEP) to seek expert input on the Development and Maintenance of Quality Measures for the Long-Term Care Hospital Quality Reporting Program (LTCH QRP). This all-day, in-person TEP meeting was held on March 28, 2017 in Baltimore, MD. The objectives of the TEP meeting were to obtain input on current LTCH QRP quality and resource use measures implemented in the program and obtain guidance and recommendations for future measures. The following experts participated in this TEP:

Susan Bowen, RN, CCRN, CPHQ, CLNC, Director, Quality/Outcomes/Patient Safety Officer at

Shepherd Center

Jean M. de Leon, MD, FAPWCA, Professor; Physical Medicine & Rehabilitation; Medical Director of Wound Care at University of Texas Southwestern Medical Center

Karen Finerty, RN, BSN, MBA, Chief Quality Officer at RML Specialty Hospital

Meg Hassenpflug, MS, RD, FCCM, Director of Outcomes and Value at Barlow Respiratory Hospital

James Jewell, MD, Medical Director, Medical Acute Care Unit at Hebrew Rehabilitation Center

Steven Lichtman, EdD, MAACVPR, Patient representative, Director, Cardiopulmonary Outpatient Services, Rehabilitation Research; Research Scientist at Helen Hayes Hospital

Sean Muldoon, MD, MPH, MS, Corporate Senior Vice President and Divisional Chief Medical Officer at Kindred Healthcare

William J. Reilly, MS OTR/L Director of Inpatient Rehabilitation at Spaulding Hospital for Continuing Care

Mary Van de Kamp, MS/CCC-SLP, Senior Vice President of Quality at Kindred Healthcare

John Votto, DO, FCCP, Executive Liaison at Hospital for Special Care and Professor of Clinical Medicine at University of Connecticut, School of Medicine and Yale University School of Medicine

Previous TEP meetings:

The first expert panel meeting, held as part of a project titled Analysis of Crosscutting Medicare Quality Metrics Using the Uniform Assessment Tool Developed and Tested as Part of the CMS Post-Acute Care Payment Reform Demonstration, was funded by the Assistant Secretary for Planning and Evaluation. The expert panel meeting was held on August 15, 2012, in Washington, DC, with the following expert panel members:

James Farrell, CNO, Healthsouth

David Gifford, MD, MPH, Senior Vice President for Quality & Regulatory Affairs at American Health Care Association

Eileen Bach, PT, MEd, DPT, Compliance Specialist, Director Quality and Patient Safety at Visiting Nurse Service of New York

Linda Resnik, PhD, PT, Associate Professor of Health Services, Policy and Practice at Brown University

Trudy Mallinson, PhD, OT, Assistant Professor at University of Southern California, Department of Occupational Science and Occupational Therapy

Margaret Stineman, MD, Professor of Physical Medicine and Rehabilitation, Vice Chair & Director, Research, Department of Physical Medicine & Rehabilitation at University of Pennsylvania

Margaret Rogers, PhD, Chief Staff Officer for Science & Research at American Speech-Language-Hearing Association

Pam Roberts, PhD, OTR/L, CPHQ, FAOTA, Manager at Cedars-Sinai Medical Center

Bruce Gans, MD, Executive Vice President and Chief Medical Officer at Kessler Institute

William Pesce, DO, Chief of Physical Medicine & Rehabilitation at Hospital for Special Care

Roger Herr, PT, MPA, COS-C, Vice President Quality Management at Independence Care System

A second expert panel meeting was held on April 15, 2013, as part of a project entitled Symptom Management Measure Development. The following LTCH experts were included on this panel:

Alfred Chiplin, JD, Senior Policy Attorney at Center for Medicare Advocacy

Dexanne Clohan, MD, Senior Vice President and Chief Medical Officer at HealthSouth

Margaret Crane, RN, CEO at Barlow Respiratory Hospital

Jean M de Leon, MD, Medical Director Wound Care at Baylor Specialty Hospital

Thomas Durkin, MHA, CRRN, RN, Executive Vice President at Vibra Healthcare

Maura A. Hopkins, RN, MSN, NEA-BC, Vice President, Patient Care Services/Chief Nursing Officer at RML Specialty Hospital

Gary Kempf, RN, Chief Clinical Executive at Christus Dubuis Health System

Dana Mukamel, PhD, Professor in the Department of Medicine at the Health Policy Research Institute at the University of California, Irvine

Sean Muldoon, MD, MPH, Chief Medical Officer at Kindred Healthcare

Terrence O'Malley, MD, Medical Director, Non-Acute Care Services for Partners Healthcare

Lisa Snyder, MD, MPH, Chief Quality Officer at Select Medical Corporation

Sharon Sprenger, MPA, RHIA, CPHQ, Senior Advisor, Measurement Outreach, Division of Healthcare Quality Evaluation at The Joint Commission

Patricia M. Stimac, MS, RD, LDN, NHA, Director of Quality Management, Nursing Home Administrator Director of Nutrition at Spartanburg Hospital for Restorative Care

John J. Votto, DO, President and CEO at Hospital for Special Care

A third expert panel meeting was held in Baltimore, MD, on September 9, 2013, as part of a project titled Symptom Management Measures. The following experts served on this panel:

Lawrence Miller, MD, Clinical Professor of Medicine at the University of California, Los Angeles

Richard Black, MD, Corporate Rehabilitation Consultant at HCR Manor Care

Mary Van de Kamp, MS, CCC-SLP, Senior Vice President of Quality and Care Management at Kindred

Timothy Reistetter, PhD, OTR, Associate Professor at University of Texas Medical Branch

Ellen Strunk, PT, MS, GCS, Consultant at Rehab Resources & Consulting, Inc.

Saad Naaman, MD, MS, Clinician at Physiatry (Physical Medicine & Rehabilitation) Practice

Linda Ladesich, MD, Medical Director Sunflower State Health

Paulette Niewczyk, MPH, PhD, Director of Research at the Uniform Data System for Medical Rehabilitation

Camille Haycock, RN, MS, Vice President, Care Continuum at Catholic Health Initiatives

Elizabeth Newman, OTD, OT/L, Director of Occupational Therapy, Rehabilitation Engineering, and Clinical Informatics at Medstar National Rehabilitation Hospital

Karon Cook, PhD, Research Associate Professor at Northwestern University

Richard Riggs, MD, Chairman and Medical Director for Cedars-Sinai Medical Center

Michelle Camicia, MSN, RN, Director of Operations at Kaiser Foundation Rehabilitation Center

Jill Bolte Taylor, PhD, Author, My Stroke of Insight

Measure Developer/Steward Updates and Ongoing Maintenance

Ad.2 Year the measure was first released: 2015

Ad.3 Month and Year of most recent revision: 10, 2017

Ad.4 What is your frequency for review/update of this measure? annually

Ad.5 When is the next scheduled review/update for this measure? 04, 2019

Ad.6 Copyright statement: Not applicable

Ad.7 Disclaimers: Not applicable

Ad.8 Additional Information/Comments: Not applicable