NATIONAL QUALITY FORUM
Driving measurable health improvements together

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: Ctrl + click link to go to the link; ALT + LEFT ARROW to return

Brief Measure Information

NQF #: 0689
Corresponding Measures:
Measure Title: Percent of Residents Who Lose Too Much Weight (Long-Stay)
Measure Steward: Centers for Medicare & Medicaid Services
sp.02. Brief Description of Measure: This measure captures the percentage of long-stay nursing home residents with a target Minimum Data Set (MDS) 3.0 assessment (OBRA, PPS, or discharge) that indicates a weight loss of 5% or more of the baseline weight in the last 30 days, or 10% or more of the baseline weight in the last 6 months, which is not a result of a physician-prescribed weight-loss regimen. The baseline weight is the resident's weight closest to 30 or 180 days before the date of the target assessment. Long-stay nursing facility residents are identified as those who have had 101 or more cumulative days of nursing facility care.
1b.01. Developer Rationale: This outcome-based quality measure the percentage of long-stay nursing home residents who experience a weight loss of 5% or more of the baseline weight in the last 30 days, or 10% or more of the baseline weight in the last 6 months, which is not a result of a physician-prescribed weight-loss regimen. Unintended and excessive weight loss is important to monitor in the nursing home population because of the impact on health outcomes, as weight loss is associated with higher risk of hospitalization and increased mortality (Xu et al, 2019; Keller et al., 2014; Stack et al., 2013; Söderström et al., 2017; Wirth et al., 2016; Wirth et al., 2018). Furthermore, evidence suggests that weight loss is associated with increased mortality in the elderly population regardless of baseline BMI (Pizzato et al., 2015). Additionally, studies have shown that risk for unintended weight loss is associated with a variety of resident characteristics, including but not limited to, increasing age, low BMI, eating dependency, absence of teeth or dentures, depression, severe cognitive impairment, low functional status, and dementia (Torbahn et al, 2021; Sanford et al., 2020; Beck, A. M., 2015; Keller et al., 2017; Wirth et al., 2018; Velázquez-Alva et al., 2020; Madeira et al., 2019; de Souto Barreto et al., 2017). The capacity of nursing homes to provide residents sufficient support, such as feeding assistance and nutritional supplements, by appropriately trained nursing and non-nursing staff to serve the medical needs of residents can mitigate nutritional risks and prevent unintended weight loss.

References:
Beck, A.M. (2015). Weight loss, mortality and associated potentially modifiable nutritional risk factors among nursing home residents — A Danish follow-up study. Journal of Nutrition, Health and Aging; 96–101. https://doi.org/10.1007/s12603-015-0439-6
de Souto Barreto, P., Cadroy, Y., Kelaiditi, E., Vellas, B., Rolland, Y. (2017). The prognostic value of body-mass index on mortality in older adults with dementia living in nursing homes. Clinical Nutrition; 36(2), 423-428. doi: 10.1016/j.clnu.2015.12.009. Epub 2015 Dec 18. PMID: 26724185.
Keller, H., Beck, A. M., & Namasivayam, A. (2014). Improving food and fluid Intake for older adults living in long-term care: A research agenda. Journal of the American Medical Directors Association.

Madeira, T., Peixoto-Plácido, C., Sousa-Santos, N., Santos, O., Alarcão, V., Goulão, B., . . . Gorjão Clara, J. (2019). Malnutrition among older adults living in Portuguese nursing homes: The PEN-3S study. Public Health Nutrition, 22(3), 486-497. doi:10.1017/S1368980018002318
Pizzato, S., Sergi, G., Bolzetta, F., De Rui, M., De Ronch, I., Carraro, S., Berton, L., Orr, E., Imoscopi, A., Perissinotto, E., Coin, A., Manzato, E., Veronese, N. (2015). Effect of weight loss on mortality in overweight and obese nursing home residents during a 5-year follow-up. European Journal of Clinical Nutrition; 69(10):1113-8. doi: 10.1038/ejcn.2015.19. Epub 2015 Mar 11. PMID: 25758838
Sanford, A.M., Morley, J.E., Berg-Weger, M., Lundy, J., Little, M.O., Leonard, K., Malmstrom, T.K. (2020). High prevalence of geriatric syndromes in older adults. PLoS ONE 15(6): e0233857. https://doi.org/10.1371/journal.pone.0233857
Söderström, L., Rosenblad, A., Thors Adolfsson, E., & Bergkvist, L. (2017). Malnutrition is associated with increased mortality in older adults regardless of the cause of death. British Journal of Nutrition,117(4), 532-540. doi:10.1017/S0007114517000435
Stack, S., Chertow, G. M., Johansen, K. L., Si, Y., & Tamura, M. K. (2013). Pre-ESRD changes in body weight and survival in nursing home residents starting dialysis. Clinical Journal of the American Society of Nephrology, 8(10), 1734-1740.
Torbahn, G., Sulz, I., Großhauser, F., Hiesmayr, M.J., Kiesswetter, E., Schindler, K., Sieber, C.C., Visser, M., Weber, J., Volkert, D. (2021). Predictors of incident malnutrition-a nutritionDay analysis in 11,923 nursing home residents. European Journal of Clinical Nutrition. doi: 10.1038/s41430-021-00964-9. Epub ahead of print. PMID: 34239065.
Velázquez-Alva, M.C., Irigoyen-Camacho, M.E., Cabrer-Rosales, M.F., Lazarevich, I., Arrieta-Cruz, I., Gutiérrez-Juárez, R., Zepeda-Zepeda, M.A. (2020). Prevalence of Malnutrition and Depression in Older Adults Living in Nursing Homes in Mexico City. Nutrients; 12(8), 2429. https://doi.org/10.3390/nu12082429
Wirth, R., Streicher, M., Smoliner, C., Kolb, C., Hiesmayr, M., Thiem, U., Sieber, CC., Volkert, D. (2016). The impact of weight loss and low BMI on mortality of nursing home residents - Results from the nutritionDay in nursing homes. Clinical Nutrition: 35(4), 900-6. doi: 10.1016/j.clnu.2015.06.003. Epub 2015 Jun 19. PMID: 26143743.
Wirth, R., Pourhassan, M., Streicher, M., Hiesmayr, M., Schindler, K., Sieber, C.C., Volkert, D. (2018). The Impact of Dysphagia on Mortality of Nursing Home Residents: Results From the nutritionDay Project. Journal of the American Medical Directors Association; 19(9):775-778. doi: 10.1016/j.jamda.2018.03.016. Epub 2018 May 31. PMID: 29778638.
Xu, D., Kane, R., Arling, G. (2019). Relationship between nursing home quality indicators and potentially preventable hospitalisation. BMJ Quality & Safety; 28(7):524-533. doi: 10.1136/bmjqs-2018-008924. Epub 2019 Mar 13. PMID: 30867234.

---

sp.12. Numerator Statement: The numerator is the number of long-stay nursing home residents with a selected target assessment indicating a weight loss of 5% or more of the baseline weight in the last 30 days or 10% or more of the baseline weight in the last 6 months who were not on a physician-prescribed weight-loss regimen (K0300 = [2]). The baseline weight is the resident's weight closest to 30 or 180 days before the date of the target assessment.

sp.14. Denominator Statement: The denominator includes all long-stay residents in the nursing home who have a target assessment (OBRA, PPS or discharge) during the selected quarter and who do not meet the exclusion criteria.

sp.16. Denominator Exclusions: There are four exclusions applied to the denominator: (1) the target assessment is an OBRA admission assessment (A0310A = [01]) or a PPS 5-day assessment (A0310B = [01]), (2) having a prognosis of life expectancy of less than six months (J1400 = [1]) or the six-month prognosis item is missing (J1400 = [-]) on the target assessment, (3) receiving hospice care (O0100K2 = [1]) or the hospice care item is missing (O0100K2 = [-]) on the target assessment, or/and (4) the weight loss item is missing (K0300 = [-]) on the target assessment. Only 1,551 episodes in the 2019 (Q1-Q4) long stay resident sample were excluded from the denominator for this measure due to missing responses on the prognosis of life expectancy being less than 6 months, which accounts for 0.04% of the total episodes. Additionally, only 7,948 (0.241%) episodes in

the 2019 (Q1-Q4) long stay residents sample were excluded due to missing responses for the Hospice care item, and only 30,854 (0.935%) episodes were excluded due to missing responses for the weight loss item. If the facility sample includes fewer than 20 residents after all other resident-level exclusions are applied, then the facility is suppressed from public reporting because of small sample size.

Measure Type: Outcome
sp.28. Data Source: Assessment Data
sp.07. Level of Analysis: Facility

IF Endorsement Maintenance – Original Endorsement Date: 2011-03-03
Most Recent Endorsement Date: 12/10/2015

IF this measure is included in a composite, NQF Composite#/title:
IF this measure is paired/grouped, NQF#/title:
sp.03. IF PAIRED/GROUPED, what is the reason this measure must be reported with other measures to appropriately interpret results?:

Preliminary Analysis: Maintenance of Endorsement Measure

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.

Summary of prior review in 2015

- This is an outcome measure utilizing Omnibus Budget Reconciliation Act (OBRA), Prospective Payment System (PPS) or discharge assessment data at the facility level to assess the number of long-stay nursing home residents with a selected target assessment indicating a weight loss of 5% or more of the baseline weight in the last 30 days or 10% or more of the baseline weight in the last 6 months who were not on a physician-prescribed weight-loss regimen.
- The logic model presented by the developer for this outcome measure links actions that can be taken by the accountable entity— such as staff education and training, adherence to guidelines, and number

of staff with rates of weight loss not a result of a prescribed regimen as well as rates of hospitalization, ED use, and mortality.

- Prior evidence submitted by the developer demonstrated that weight loss of an excessive an unintended nature is problematic for long-stay nursing home residents. Research cited by the developer revealed an association between unintended weight loss and the structures and care processes in the nursing home, thereby reflecting a clear opportunity to analyze quality of care.
- The developer referenced a number of structural and process changes that can improve residents' nutritional status: higher staffing levels, communication training, and changes to the dining environment.
- During the last review cycle, the Standing Committee agreed that the evidence was strong for this important outcome measure. They raised concerns with the lack of data on disparities and the lack of observable improvements. The developers stated that the lack of change in this measure may indicate that nursing homes are not improving in this area, highlighting the need for continued public reporting on it. It was also noted by the Committee that as there is a greater effort to keep people at home as long as possible. The population in nursing homes is increasingly frail, which leads to difficulty in maintaining nutritional status.

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:

- The developer offered a number of updated research studies to further demonstrate the negative association between excessive or unintentional weight loss and other health outcomes, including risk of hospitalization and increased mortality, lower BMI, lower masticatory function, and residents with behavioral health or neurological disorders such as depression and dementia.
- The developer also presents new evidence of the impact of COVID-19 on both residents who contracted the virus and those who did not. Small percentages of residents experienced unintended weight loss in both infected and non-infected populations (10 and 7.5% respectively) in the study from 2021.
- The developer also presented updated evidence on several actions that nursing home staff and facilities can take to prevent unintended weight loss.

Questions for the Committee:

- *The evidence provided by the developer is updated and directionally the same 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 a health outcome (Box 1) → Relationship between measured health outcome and healthcare action is demonstrated (Box 2) → 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.

- The developer reported an analysis of four-quarter facility-level data (14,274 facilities between Q1 and Q4 of 2019).
    - The mean performance was 5.2 percent with a standard deviation of 3.1 percent and range of 1.6 percent to 9.2 percent.
    - The developer also notes that the IQR of 3.9 percent, and the small number of facilities with "perfect" scores (2.6%) indicates room for improvement.

Disparities

- The developer examined the same data for associations between measure scores and age, race, and Medicaid enrollment.
    - Population older than 85 years is at slightly higher risk of losing too much weight (5.9%) when compared to younger residents (5.1%) ($p \leq 0.0001$)
    - White population is at slightly higher risk of losing too much weight (5.4%) when compared to non-white population (5.2%) ($p \leq 0.0001$)
    - Non-Medicaid enrollees are at slightly higher risk of losing too much weight (6.27%) when compared to Medicaid enrollees (5.20%) ($p \leq 0.0001$)

Questions for the Committee:

- Is there a gap in care that warrants a national performance measure?
- Are Standing Committee members aware of any other potential disparities between long-stay nursing home residents that would warrant a national measure?

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)

- Since the scientific methods committee, is use agnostic, it is the responsibility of this committee to evaluate the evidence and context carefully. The evidence that care at a LTC facility is associated with weight loss is mentioned but not shared. Evidence that weight loss is bad and that there is variability in weight loss is presented. The major issue is confounding. Of course weight loss is bad outcome. A patient with a new stroke, malignancy or many other health conditions are associated with bad outcomes and LTC placement. Nor was there evidence presented that resident weight loss is a modifiable outcome. This was addressed partly in the validity section for risk adjustment, however this issue of causality is important enough to be addressed in the very beginning with evidence
- This is a maintenance measure, and the developer provided updated information linking weight loss and adverse outcomes, as well as weaker evidence on care quality and weight loss.
- Evidence to support this outcome measure is strong; several processes (adequate time, more staff, nutritional supplements, family involvement, etc. can impact the outcome). This intermediate outcomes, impacts other long term outcomes: re-hospitalization and mortality.
- Strong face validity for this measure, but less specific evidence provided by developer.
- outcome measure - adequately supported
- adequate update on evidence provided
- We cannot ignore recent studies exploring the impact of COVID 19 on the current staffing crisis and on patient isolation in long term care. Critical staffing levels among healthcare workers and bans on visitation may be uncontrollable factors setting facilities up for failure.
- Evidence suggests that unintended excessive weight loss is associated with an increased morality in the elderly population and can also be an indicator of quality and safety care at the long-term nursing facilities. So, the evidence highly supports this outcome measure, which can be used to drive quality improvement for long-term care. The developer also submitted updated studies and new evidence of COVID-19 impact on residents' weight loss.
- no need for further discussion of evidence

- Clearly linked to important health status and is actionable; no evidentiary threats since last review
- Agree with "pass" as there are interventions that have been shown to address weight loss including family member feeding programs and nutritional supplements.
- I would like to see evidence that this measure has impacted outcomes since original 2011 endorsement. Evidence presented does not pertain to the impact of the measure in terms of outcomes..
- Yes. Maintenance measure that must incorporate the impact of Covid 19 as well as trends to age in home, which are noted by developer.
- Weight loss in long term nursing home residents is associated with increased morbidity and mortality. This is directly measured and objective.
- ok
- The evidence presented for this measure is directly related to the measure, and they have parted it with improvement actions that could be levers to effect poor performance. They have also presented new evidence that supports the measure's focus.
- Variability in performance cannot adequately be assessed unless the groups are equivalent and we don't know that. Behavioral health care centers with a high incidence of depression are not comparable to other basically well resident facilities
- small range in this outcome which is (fortunately) somewhat rare. However, enough variation between facilities exists to support a performance gap. Very few facilities have perfect scores.
- There is variability in performance across 4 quarters in 2019. 10th percentile: 1.6%, Median 4.8%. However, data has not really changed over time. Disparities were assessed across age (<85y vs. >=85y), Race/ethnicity (whites at higher risk), Medicaid enrollment (non-Medicaid enrollees at higher risk)
- Large performance gap in facilities, though this is likely to substantially drop with new supreme court ruling on facilities that receive CMS funding. Disparities data presented for geographic location and by racial/ethnic identity
- Challenges with disparities but comparisons between medicaid and non medicaid, white vs non white and age were conducted
- information provided on both
- A gap in care appears to remain. But does this support the continued need for this measure? Difficult to say when we can't control for the intrinsic motivation of an aged adult to thrive.
- An analysis of performance data from 14,274 facilities were provided for the period of Q1 and Q4 of 2019. The range of performance is between 1.6% and 9.2%, indicating a performance gap. The opportunity for improvement was rated as “moderate.” Disparities were also examined based on age, race/ethnicity, gender, socioeconomic status, etc. The developer found residents older than 85 years, whites, and non-Medicaid enrollees have a slightly higher risk of losing too much weight when compared to younger residents, non-white population, and Medicaid enrollees.
- There is a performance gap that warrants measurement. Disparities noted.
- There is clear variation and only limited sites with very high performance; disparity/subpopulation data still lag
- "Moderate" as the range is 1.6-9.2% I think there’s enough of a variation to warrant measure to help facilitate improvement
- Mean performance 5.2 percent with SD 3.1 percent. The committee could discuss whether this warrants a national performance measure. Has this improved since 2011?
- yes
- This is an ongoing measure. Quality of care has improved since the measure was introduced.
- interesting disparity results
- There does appear to be a performance gap that warrants a national measure based on the IQR, however the disparities data is not compelling.
- No concerns based on initial assembly or ongoing use, with appropriate mathematic handling

- no concerns
- AL
- ok

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:
Alex Sox-Harris, Samuel Simon, Zhenqiu Lin, Laurent Glance, Matt Austin, Terri Warholak, Jeffrey Geppert, Christie Teigland, Eugene Nuccio, Lacy Fabian, Marybeth Farquhar, Joseph Kunisch

Methods Panel Review (Combined)

Methods Panel Evaluation Summary:
This measure was reviewed by the Scientific Methods Panel and discussed on the call. A summary of the measure and the Panel discussion is provided below.

Reliability

- Patient/Encounter Level (i.e. Data Element) Testing
    - The 2015 submission included data element testing was conducted from August 2006 to February 2007 using the MDS 3.0. The developer reports that no changes have been made to the specific MDS items used in this measure since the testing was last conducted.
    - The RAND Development and Validation of MDS 3.0 study sample included 3,822 residents from 71 community nursing homes and 764 residents from 19 VHA nursing homes (Saliba & Buchanan, 2008).
    - In their testing of the MDS 3.0, RAND observed the kappa assessments as follows:
        - The kappa for gold-standard to gold-standard assessments of the weight loss item was 0.944

        - The kappa for gold-standard nurse assessment to facility nurse assessment of weight loss item was 0.918.
        - The kappa for gold-standard to gold-standard assessments of the 6-month prognosis item was 0.8724.
        - The kappa for gold-standard nurse assessment to facility nurse assessment of the 6-month prognosis item was 0.964.
- Accountable Entity Level (i.e. Performance Measure Score) Testing
    - The developer conducted performance score/accountable entity level testing for the 2021 submission using 2019 data from the MDS 3.0
    - The developer's national test of MDS 3.0 items used data from all long-stay residents who met the denominator inclusion criteria for this measure in facilities with sufficient sample size ($n \geq 20$, $k = 14{,}274$) and reported this measure between 2019Q1 and 2019Q4. 932,094 residents met the denominator inclusion criteria for testing in these facilities.
    - The developer used inter-rater reliability testing to examine the agreement between assessors. Inter-rater reliability measures the extent to which two data collectors achieve the same results when assessing the same resident within the same time frame. The developer presented two types of reliability: gold-standard nurse to gold-standard nurse, and gold-standard nurse to facility-nurse.
    - The developer also conducted split-half reliability testing and signal-to-noise testing, reporting that the split-half correlation for this measure was positive, and the relationship was moderate ($r = 0.64$, $\rho = 0.65$, $ICC = 0.64$, $p < .01$), suggesting there is evidence of internal reliability.
    - The developer reported an average signal-to-noise reliability score of 0.76 using facility scores based on 2019Q1-2019Q4 data.

Validity

- Patient/Encounter Level (i.e. Data-Element) Testing
    - In the 2015 submission, the developer reported the kappa for gold-standard nurse assessment to facility nurse assessment of weight loss item was 0.918. The kappa for gold-standard nurse assessment to facility nurse assessment of the 6-month prognosis item was 0.964.
- Accountable Entity Level (i.e. Performance Measure Score) Validity
    - The developer hypothesized a number of relationships to test the measure. Tests of convergent validity, variation by state, seasonality, stability analysis, and confidence interval analysis were run to demonstrate validity of the measure.
        - Convergent Validity:
            - Among facilities that could report all measures, the analysis found statistically significant positive correlation between Percent of Residents Who Lose Too Much Weight (NQF #0689) and the following: Percent of Residents Whose Ability to Move Independently Worsened (0.113), Percent of Residents Whose Need for Help with Activities of Daily Living Has Increased (0.108), and Percent of Residents Who Have Depressive Symptoms (0.063).
            - The developer reported statistically significant negative correlations between Percent of Residents Who Lose Too Much Weight and the following as well: Overall Facility Five-Star Ratings (0.108), Quality Ratings (0.143), Staffing Ratings (0.029), and Registered Nurse Staffing Ratings (0.011) were also observed.
        - Variation by State:

- The developer hypothesized that "if a measure is subject to variation caused by other factors beyond facility control, such as state-level payment policies or demographics, this variation can be a threat to the validity of the measure. At the same time, it is expected that state variation may explain a small portion of measure variation due to differences in quality across states."
- The developer found that the proportion of variation in this measure explained by the state that facilities are located in was small though significant ($p < .001$).
- An analysis of variance showed that 4.8% of the overall variance in this measure can be attributed to the state in which the facility is located. The average IQR of state-level scores was 3.8 percentage points.
- Alaska had the highest mean for NQF #0689 compared to the other states but had a relatively low number of nursing facilities located in the state.

- Seasonality: the developer presented a graph of the data and concluded that only "slight" seasonal variation was seen. However, no empirical results were presented.
- Stability Analysis: the developer reported that 39.2 percent did not change deciles, 37.9 percent changed one decile, and 22.9 percent changed two or more deciles
- Confidence interval analysis: the developer reported that 28.7 percent of facilities had a significantly different score from the national mean (15.4% were lower and 13.2% were higher).

- Although the hypothesized relationships were demonstrated by the developer to be in the expected direction, some SMP members noted that the correlations were not strong enough to demonstrate validity. Other SMP members raised concerns with the method, rather than their results, stating that convergent validity should not have been used to demonstrate validity of the measure, resulting in weak correlations.
- Multiple SMP members raised concerns about the developer's decision not to risk adjust the measure. The SMP suggested that the Standing Committee should discuss whether certain MDS score items (e.g. for depression, Alzheimer's, cancer, Parkinson's, cognitive impairment, cardiac disorders, benign gastro diseases, eating dependencies, and swallowing/chewing problems) warrant an examination of risk adjustment.
- In response to these SMP concerns, the developer conducted testing of two risk models, which are available in the submission appendix. The developer concluded that risk adjustment was not appropriate for the measure based on the conceptual model and the risk models' limited impact on provider scores.
- Overall, consensus was not reached on the validity of this measure during the SMP measure evaluation meeting.

Questions for the Committee regarding reliability:

- Does the committee have concerns about the results of the reliability testing?
- *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:

- Does the committee have concerns with the results of the validity testing, specifically with the strength of the convergent validity results and the lack of risk adjustment for observable factors that would impact weight loss?

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

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

* SMP did not reach consensus on validity, therefore no preliminary rating is given

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

- The reliability is reasonable - weight is easily measured and likely to be reliable
- Reliability was strong against gold standards, nursing assessments and split sample testing. No major concerns, Agree with SMP rating.
- None.
- Moderate reliability, but I think there are some concerns about a 1 week look back period given the number of part time and per-diem staff that work in facilities. I am also concerned about this measure when considering that contract employees are an optional reporting category which means facilities with good numbers will report and potentially mask full time staff rates.
- Scientific panel reviewed and passed Moderate reliability
- no concerns, given performance history
- None
- Reliability and Validity were evaluated by the Scientific Method Panel. They rated reliability as moderate but did not reach consensus on validity.
- Reliability results are adequate.
- The ongoing assessments of reliability look solid and no real concerns
- Agree with scientific panel "Moderate"
- moderate level reliability
- Clearly specified
- I have no concerns about implementation of this measure.
- ok
- No concerns
- no
- No
- none; Internal consistency, split-half reliability r=0.64
- Yes, moderate concerns given the optional reporting for contract staff
- no
- no
- None
- No concerns.
- no
- No
- Agree with scientific panel "Moderate"
- moderate level reliability
- no
- no
- no concerns
- The data elements are clearly defined as evidence by the patient/encounter level testing and the high kappa. The IRR however at the to the accountable entity were significantly lower however with only moderate performance related to internal reliability.
- Yes. Confounding and comparing very different groups of people in different facilities makes this metric not valid. If we understood all of the variables involved, risk adjustment may be a approach but I am not we understand all of the mechanisms of weight loss to be able to create an appropriate model

- Yes, there are several concerns brought up by the SMP that warrant comment. There is a lot of change over time across facilities, with several facilities (22%) changing two or more deciles. This might suggest patient level, vs facility level variation and raises concern for facilities that care for complex patients. This concern is magnified by low correlations with staffing measures.
- Convergent validity: positive correlation between resident weight loss & decrease in ability to move independently (0.113), increased need for ADL assistance (0.108), and depression (0.063)--expected direction, however low correlation.
- no concerns.
- Lack of consensus by panel
- no
- No
- No concerns.
- Risk adjustment may need further discussion.
- Some weakness in the granular links, esp among subpopulations - notably seasonal threats, non-physician and yet appropriate weight efforts, and others.
- There was question about risk adjustment. I don't think there should be risk adjustment for diseases, indeed many of the literature on evidence that there are modifiable risk factors/actionable items were specific to some of these disease such as Alzheimer's. I think excluding those at the end of life is sufficient. I think Moderate on validity
- Overall consensus not reached during the SMP measure eval period. We should discuss.
- No concerns
- no
- no concerns
- That it did not include risk adjustment
- Please see above. Risk adjustment only works if we understand the causal mechanism and the factors involved in the outcome and we have not seen enough evidence about this. The correlation between other conditions like depression and increased mortality and neurologic disorders seem to suggest that causality is reversed
- I have strong concerns about the lack of risk-adjustment. Complex patients like those with dementia often are clustered within specific facilities, which may be unfairly penalized for weight loss among patients. There are several other clinical populations for whom this may be a concern as well, and warrants further investigation.
- This outcome measure is not risk adjusted. Higher social risk factors include age and race. Older avg. age of white residents may account for this. Should this be risk-adjusted (based on location, facilities may have higher proportion of white or residents >=85y)?
- Exclusions appear appropriate. Measure not risk adjusted.
- Lack of risk adjustment
- appropriate
- Measure is not risk adjusted. Explanation offered in Appendix B.
- The SMP raised concerns about whether there should be a risk adjustment for certain conditions, such as depression, Alzheimer's, cancer, Parkinson's, cognitive impairment, cardiac disorders, benign gastro diseases, eating dependencies, and swallowing/chewing problems, etc. In response to the concerns, the developer conducted testing with two risk models. They concluded that "risk adjustment was not appropriate for the measure based on the conceptual model and the risk models' limited impact on provider scores."
- Risk adjustment needs attention.
- Another weaker facet - see above re: subgroups and other threats that could influence observations

- I don't think there's rationale for risk adjustment, yes many diseases associated with weight loss increase with age but there's no data suggesting that the interventions to improve this measure can't work in the older vs. younger people so even if some facilities have older residents I'd think they should be able to use the same interventions to remedy the situation. I worry risk adjustment by age will create more ageist bias - that we'll tolerate more issues just because someone is older
- No concerns except exclusion of risk adjustment as some MDS do indicate a further examination for inclusion.
- ok
- Risk adjustment should be considered
- No Concerns

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.

- The developer attests that all data is generated during provision of care, and all data elements are in defined fields in electronic clinical data.
- The measure relies on data from the Minimum Data Set (MDS) 3.0, which the developer reports is "mandatory for all Medicare/Medicaid certified nursing facilities."
- The developer reports that 1.216 percent of data were missing from episodes in 2019, and these were excluded from the denominator. The missingness does not warrant concern for the developer as to the feasibility or bias of the measure.

Questions for the Committee:

- Does the Standing Committee have any concerns about the feasibility of using the data collection instrument (MDS 3.0)?

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

Committee Pre-evaluation Comments:
Criteria 3: Feasibility

- this is feasible
- Comes from MDS measures, no concerns
- No concerns. Mandatory reporting for CMS certified nursing facilities.
- Moderate feasibility though takes significant record review to document exclusions. Are individual facilities documenting the same way, using same criteria to apply religious, medical, or other reasons for not getting vaccinated.
- High
- no concerns given history
- None
- The developer states that the measure relies on data from the Minimum Data Set (MDS) 3.0, which report is mandatory for all Medicare/Medicaid certified nursing facilities. The test using 2019 data does not suggest concerns for feasibility or bias of the measure. The feasibility is rated as high.
- no concerns
- No concerns
- High feasibility
- Electronic clinical data.
- No concerns
- This is a basic measure that is routinely generated.
- ok
- No concerns about 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**

Accountability program details

- The measure is used for accountability as it is publicly reported. The developer does not report plans to use the measure in other accountability programs.
- The developer notes that the measure is publicly reported to both measured facilities and the public via the following:
    - Care Compare
    - Provider Data Catalog

- Certification and Survey Provider Enhanced Reports (CASPER)

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

- The developer states that measure users are given an opportunity to view performance results prior to public reporting. This data is supplemented by state and national averages, and comparisons to the averages.
- The developer states that measure users are given an opportunity to identify errors in data
- A Help Line is available to measure users for their questions

Additional Feedback:

- The measure developer analyzed inquiries submitted to their support inbox. They report that they did not receive any feedback or concerns from those being measured, measure users, or implementers since October 2019.

***Questions for the Committee**:*

- How has the measure been vetted in real-world settings by those being measured or others?
- Does the measure developer offer adequate opportunities to measure users to provide feedback on the measure's functioning?

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

- The developer stated that some change in performance was observed from 2011 – 2019. Between those years, the facility-level mean and median scores decreased.
- The developer reported the following scores:
  - 2011 mean = 6.4 percent
  - 2011 median = 5.8 percent.
  - 2018 median = 4.55 percent.
  - 2019 mean = 5.44 percent
  - 2019 median 4.76 percent.
- The developer noted that the decreasing scores over time demonstrate an improvement in the quality of care.

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

- The developer reported that after consulting the literature, it was unexpected that White residents were at greater risk of unintended weight loss than non-write residents.

- The developer conducted testing to assess whether this was due to differences in quality of care, chance, or another explanatory factor (i.e. the developer conducted risk adjustment testing). The developer concluded that another factor explained the difference, noting that "white residents being more likely to experience weight loss that is not due to a physician-prescribed weight loss regimen than non-white residents is due to the higher average age of white residents."

Potential harms

- The developer did not report any unexpected harms

Additional Feedback:

- The developer did not present additional feedback.
- This measure was not reviewed by the MAP

***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: ☒ **High** ☐ **Moderate** ☐ **Low** ☐ **Insufficient**

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

- yes
- Publicly reported, in accountability programs
- Use: Quality improvement and public reporting to Care compare
- publicly reported measure.
- Passes for use
- no concerns given history
- Yes
- The measure has been used for accountability through public reporting, such as Care Compare, Provider Data Catalog, and Certification and Survey Provider Enhanced Reports (CASPER). The measure users are given opportunities to view performance results before public reporting and to identify errors in data. A Helpline is also available for measurement users.
- Seems well vetted through feedback.
- Both issues well addressed - no concerns, being used
- High usability
- Publicly reported. Developer reports mean score improvement from 6.4 to 4.8%.
- No concerns
- The measure is publicly reported.
- ok
- No concern about use/usability
- the measure is usable if it was valid
- no unintended consequences anticipated
- no concerns. Racial differences were explained by age (confounding)
- No unintended consequences, and this is hugely important for public health
- passes - Decreasing over time from 6.4% to 4.76%
- no concerns given history
- NA
- Data provided by the developer show that the mean score was improved from 6.4 in 2011 to 4.8 in 2019. Usability is rated as high.
- Prevention of unexplained weight loss is a key part of quality healthcare. I see little opportunity for any harm from the measure.
- None of consequence noted
- Some improvement in scores from 2011-2018
- Publicly reported. Minimal harms
- No concerns
- actions to improve nutrition in the frail elderly have been implemented
- ok

Criterion 5: Related and Competing Measures

Related or competing measures
There are no related or competing measures
Harmonization
There are no related or competing measures warranting harmonization

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

- nothing relevant
- None
- No
- related measures do not appear to be a major competing concern.
- None
- NA
- No competing measures were identified.
- none indicated
- None
- None
- none
- There are none.
- no
- none
- None

Public and Member Comments

Comments and Member Support/Non-Support Submitted as of: January 19, 2022

- **No NQF members have submitted a support/non-support choice.**

Scientific Acceptability: Preliminary Analysis Form

Measure Number: *0689*
Measure Title: *Percent of Residents Who Lose Too Much Weight (Long-Stay)*
Measure is:
☐ **New** ☒ **Previously endorsed (***NOTE: Empirical validity testing is expected at time of maintenance review; if not possible, justification is required.)*

RELIABILITY: SPECIFICATIONS

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

Submission document: Items sp.01-sp.30
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.**

For example: 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?
Reviewer 3: The measure is clearly specified.
Reviewer 5: No concerns.
Reviewer 6: No concerns
Reviewer 7: Response: I do not understand statement "The baseline weight is the resident's weight closest to 30 or 180 days before the date of the target assessment." Perhaps an example would help.

Reviewer 9: The reliability and validity results for the measure score were not very good--especially for a measure that is not a new measure.
Reviewer 11: No concerns
Reviewer 12: None

RELIABILITY: TESTING

Type of measure:

☐ **Process** ☐ **Process: Appropriate Use** ☐ **Structure** ☐ **Efficiency** ☐ **Cost/Resource Use**
☒ **Outcome** ☐ **Outcome: PRO-PM** ☒ **Outcome: Intermediate Clinical Outcome** ☐ **Composite**

Data Source:

☐ **Claims** ☐ **eCQM (HQMF) implemented in EHRs** ☐ **Abstracted from Electronic Health Records**
☐ **Abstracted from Paper Medical Records** ☒ **Instrument-Based Data** ☐ **Registry**
☐ **Enrollment Data** ☒ **Other (please specify)**

Reviewer 1: Assessment data
Reviewer 3: Assessment data
Reviewer 4: Assessment data
Reviewer 5: nursing home MDS
Reviewer 6: Assessment data
Reviewer 9: nursing home minimum data set MDS 3.0
Reviewer 9: MDS 3.0
Reviewer 11: Assessment data
Reviewer 12: Assessment from clinician survey

Level of Analysis:

☒ **Group/Practice** ☐ **Individual Clinician** ☒ **Hospital/facility/agency** ☐ **Health Plan**
☐ **Population: Regional, State, Community, County or City** ☐ **Accountable Care Organization**
☐ **Integrated Delivery System** ☐ **Other (please specify)**

Submission document: Questions 2a.01-09

3. **Reliability testing level**

For example: for some types of measures, if patient/encounter level validity is demonstrated, additional reliability testing is not required. Please review table above.

☒ **Accountable-Entity Level** ☒ **Patient/Encounter Level** ☐ **Neither**

4. **Reliability testing was conducted with the data source and level of analysis indicated for this measure**

NOTE: "level of analysis" reflects which entity is being assessed or held accountable by the measure.
For example: If a measure is specified for a clinician level of analysis, but facility-level testing is provided, then testing does NOT match level of analysis. Or, if two levels of analysis are specified (e.g., clinician and facility) but testing is conducted for only one, then testing does NOT match level of analysis. Or, if claims data are selected as a data source, but testing data doesn't include claims data, then testing does NOT match data source.
Also, check "NO" if only descriptive statistics are provided or submitter only describes process for data management/cleaning/computer programming.

☒ **Yes** ☐ **No**

5. If accountable-entity level and/or patient/encounter level reliability testing was NOT conducted or if the methods used were NOT appropriate, was **empirical *VALIDITY* testing** of **patient-level data** conducted?

According to current guidance patient/encounter level validity testing can be used for patient/encounter level reliability testing. Answer ONLY if you responded "Neither" on question #3 and/or "No" to question #4. Note that for some types of measures, additional reliability testing is not required IF patient/encounter level validity is demonstrated.

☒ **Yes** ☐ **No**

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

Submission document: Question 2a.10
For example: Is the method(s) appropriate? If not, please explain (and offer potential alternatives if possible). Does the testing conform to NQF criteria and guidance? Was testing was conducted with the data source and level of analysis indicated for this measure? Address each level of testing provided, and each analysis under each method.
Reviewer 1: The methods are mostly appropriate. Its unclear if the split sample reliability was adjusted with the SB Prophesy formula. ICC for agreement would have been better than a correlation.
Reviewer 3: Data element reliability results were based on the original instrument development study. For performance score reliability, the developer conducted both split-half reliability and usual beta-binomial signal-to-noise reliability testing. All methods are appropriate.
Reviewer 4: Data reliability: Kappa analysis: gold standard nurse to facility nurse

- weight loss: 0.92
- Prognosis: 0.96

Score-level reliability: Split-half reliability testing

- ICC 0.64
- SNR: 0.76

Reviewer 5: Data element: Sailba et al. study looked at inter-rater reliability of the MDS elements (gold standard vs. gold standard; and gold standard vs local nurse) Score level: Looked at reliability in three appropriate ways - stability over time, confidence interval analysis, and signal-to-noise
Reviewer 6: No concerns
Reviewer 8: Three methods were used: 1) Stability analysis: the extent to which facility ranks and scores changed from Quarter 2 to Quarter 3 2014. Dramatic changes in the QM score or facility rank based on the score across time may indicate measure instability, rather than true changes in quality. 2) Confidence Interval Analysis: proportions of facilities with QM scores that are significantly different from the national facility-level mean, stratified by facility denominator size. A reliable measure should have a high proportion of facilities with scores significantly different than the mean. High-performing facilities should have scores that are significantly below-average, and low-performing facilities should be significantly above-average. We stratified the analysis by facility denominator size to examine whether this feature of the measure varies by size. 3) Signal to noise analysis: what proportion of total variance in the measure is attributable to differences among providers.
Reviewer 9: The submission cited a excellent study by Saliba, et al (2008) that documented the MDS 3.0 item reliability using both comparison of Gold Standard vs. Gold Standard and Gold Standard vs. Facility nurse. Additionally, for measure reliability the submission cited a study by RTI based on data
Reviewer 10: Comparison to other versions, and scores
Reviewer 11: Methods used were appropriate and conducted with the data source and level of analysis indicated for this measure.
Reviewer 12: Cited previous Rand study of the MDS 3.0 Split-half reliability analysis Signal-to-noise analysis

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

Submission document: Question 2a.11
For example: Is the test sample adequate to generalize for widespread implementation? Is there high or moderate confidence that the measure results and/or the data used in the measure are reliable? Address each level of testing provided, and each analysis under each method.
Reviewer 1: Entity-level reliability was adequate. Would have preferred to see the distribution of SNRs instead of just the average.
Reviewer 3: Split half reliability (ICC=0.64) is acceptable and average signal-to-noise reliability (0.76) is also very good. It would be helpful if the developer can provide the distribution of the signal-to-noise reliability scores across facilities. The results of data elements testing are somewhat puzzling. The kappa for gold-standard nurse assessment to facility nurse assessment of the 6-month prognosis item was 0.964 while the kappa for gold-standard to gold-standard assessment of the 6-month prognosis item was 0.872. One

would expect the opposite. This finding call into question how so-called "gold-standard" should be interpreted here. The reassuring thing is that both are high.
Reviewer 4: see above
Reviewer 5: Data element: kappas ranged between 0.87 and 0.94 on the MDS elements used in the measure Score level: stability analysis indicated nearly half of facilities saw their decile change 2+ deciles; signal-to-noise ratio was 0.078, indicating only 7.8% of variance is explained by facility characteristics.
Reviewer 6: No concerns
Reviewer 8: Data Element Reliability: The kappa for gold-standard to gold-standard assessments of the weight loss item was **0.944**; and the kappa for gold-standard nurse assessment to facility nurse assessment of weight loss item was **0.918**. The kappa for gold-standard to gold-standard assessments of the 6-month prognosis item was **0.8724**; and the kappa for gold-standard nurse assessment to facility nurse assessment of the 6-month prognosis item was **0.964. Results indicate high data element reliability**
Performance Measure Score Reliability:
Stability analysis: Change in facility ranks on weight loss quality measure between two consecutive quarters were common. More than 20% of facilities changed rank by MORE THAN 3 DECILES, 34% changed rank by 1-3 deciles, and 45% stayed within decile. Facility scores on this measure were more stable with 81.4% changing score within 1 SD and less than 1% changing score by 3+ SD. This instability is likely due to the very small variation in facility scores making this measure not very reliable.
Confidence interval analysis: Slightly more than one-third (37.4%) of nursing homes had a mean score for which the 95% confidence intervals did not overlap with the national mean; only 10.2% had scores lower than 95% CI and 26.6% had scores higher than 95% CI. Again, this measure does not show a significant number of facilities with performance scores statistically different from the national mean, indicating lower reliability. Developers unfortunately did not show the distribution of performance scores, but the difference between change in rank (OFTEN) vs. change in scores (NOT OFTEN) would indicate tight distribution.
Signal-to-noise analysis: facility was a significant predictor of QM scores ($p < 0.001$), but the **signal-to-noise ratio for this measure was low at 0.078** , indicating that **only 7.8% of the variance in scores for this measure in Quarter 1-3 2014 was explained by facility characteristics** (including the underlying quality of care in each facility). Thus, this measure is not reliable in separating facility characteristics from the noise of population variance.
Reviewer 9: The results for the data element analyses were outstanding (typically kappa > 0.90). The results for the measure score analyses were marginal. Less than ½ of the facilities (44.6%) had no quarter-to-quarter change, while 20% had a > 3 decile change quarter-to-quarter. Similarly, more than 1/3 (37.4%) had a mean score where the 95% CI did not include the national mean. The S-t-N ratio was very low at only 7.8% detectability.
Reviewer 10: Kappa: .964; .76
Reviewer 11: Test sample was adequate with moderate to high confidence that the results are reliable.
Reviewer 12: "of the MDS 3.0, RAND observed that the kappa for gold-standard to gold-standard assessments of the weight loss item was 0.944; and the kappa for gold-standard nurse assessment to facility nurse assessment of weight loss item was 0.918. The kappa for gold-standard to gold-standard assessments of the 6-month prognosis item was 0.8724; and the kappa for gold-standard nurse assessment to facility nurse assessment of the 6-month prognosis item was 0.964. Kappa is a statistical measure of inter-rater agreement for qualitative data, ranging from 0.0 to 1.0. A rating of 0.964 is considered “substantial agreement.” These results are indicative of data element reliability. 1) Split-half reliability analysis: The split-half correlation for this measure was positive, and the relationship was moderate ($r = 0.64$, $\rho = 0.65$, ICC = 0.64, $p < .01$), suggesting there is considerable evidence of internal reliability. 2) Signal-to-noise analysis: The average signal-to-noise reliability score of this quality measure using facility scores based on 2019Q1-2019Q4 data was observed to be 0.76. This suggests that the measure is very reliable in separating facility characteristics from variability within facility"

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: Question 2a.10-12

For example: Appropriate signal-to-noise analysis; random split-half correlation; other accepted method with description of how it assesses reliability of the performance score.

☒ **Yes**

☒ **No**

☐ **Not applicable**

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

Submission document: Question 2a.10-12

For example: inter-abstractor agreement (ICC, Kappa); other accepted method with description of how it assesses reliability of the data elements

Answer NO if: only assessed percent agreement; did not assess separately for all critical data elements (or at minimum, for numerator, denominator, exclusions)

☒ **Yes**

☒ **No**

☐ **Not applicable** (patient/encounter level 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** accountable-entity level testing has been conducted)

☒ **Moderate** (NOTE: Moderate is the highest eligible rating if accountable-entity 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.**

Reviewer 1: Reasonable entity-level reliability using both methods.

Reviewer 3: Performance score reliability in the range of 0.6 - 0.7 is acceptable.

Reviewer 4: Results of data and measure level reliability are acceptable.

Reviewer 5: Data element reliability demonstrated high levels of agreement between raters; score level reliability testing indicated the measure is not a reliable measure of performance.

Reviewer 7: In addition to median reliability would have been better to show reliability metrics stratified by facility volume

Reviewer 8: See explanations under evaluation of reliability testing results above to explain low rating.

Reviewer 9: The methodologies describe to assess reliability were appropriate. The results for the item reliability were excellent (high); however, the measure score reliability results were marginal at best (low). Given that the measure score is the critical value, the overall rating must be biased in that direction.

Reviewer 10: Appeared to compare the measure with prior versions of itself.

Reviewer 11: Testing at the element level and score level. Kappa scores are high.

Reviewer 12: No concerns

VALIDITY: TESTING

12. **Validity testing level (check all that apply):**

☒ **Accountable-Entity Level** ☒ **Patient or Encounter-Level** ☐ **Both**

13. **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*:* Questions 2b.01-02.
For example: Data validity/accuracy as compared to authoritative source- sensitivity, specificity, PPV, NPV; other accepted method with description of how it assesses validity of the data elements.
Answer NO if: only assessed percent agreement; did not assess separately for all critical data elements (or at minimum, for numerator, denominator, exclusions)

☒ **Yes**
☒ **No**
☐ **Not applicable** (patient/encounter level testing was not performed)

14. **Method of establishing validity at the *accountable-entity level*:**

NOTE: Empirical validity testing is expected at time of maintenance review; if not possible, justification is required.
Submission document: Questions 2b.01-02

☒ **Face validity**
☒ **Empirical validity testing at the accountable-entity level**
☐ **N/A (accountable-entity level testing not conducted)**

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

Submission document: Question 2b.02
For example: Correlation of the accountable-entity level on this measure and other performance measures; differences in performance scores between groups known to differ on quality; other accepted method with description of how it assesses validity of the performance score

☒ **Yes**
☒ **No**
☐ **Not applicable** (accountable-entity level testing was not performed)

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

Submission document: Question 2b.02
For example:

- *If face validity the only testing conducted: Was it accomplished through a systematic and transparent process, by identified experts, explicitly addresses whether performance scores resulting from the measure as specified can be used to distinguish good from poor quality, and the degree of consensus and any areas of disagreement provided/discussed?*
- *If a maintenance measure, but no empirical testing conducted, was justification provided?*
- *If construct validation conducted, was the hypothesized relationship (including strength and direction) described and does it seem reasonable?*

Reviewer 1: I would have liked to have seen a conceptual model that linked facility processes and structure to variation in this outcome, then tests of those relationships. The convergent validity analysis, although typical, is unsatisfying. The statistical significance is meaningless given the sample size. The correlations are very modest albeit in the hypothesized direction. Its unclear to me if these correlation support or fail to support the hypotheses. The state and seasonal analyses were interesting and comforting. The stability analysis is interesting but hard to interpret. Are these changes due to unreliability or real changes? The CI analysis has more to do with performance gap than validity, but there are lots of facilities that are different than the mean, which is good.
Reviewer 3: Data element validity results were based on the original development study. Empirical validity analysis, in particular convergent validity testing, was conducted to test the hypothesis of expected relationships with other related quality measures.
Reviewer 4: Examined correlation between measure and other measures:

- Percent of Residents Whose Ability to Move Independently Worsened (0.11)
- Percent of Residents Whose Need for Help with Activities of Daily Living Has Increased (0.11)
- Facility ratings – overall, quality and staffing

Reviewer 5: Data element: Relied on previous studies that have looked at inter-rater agreement Measure score: Looked at both seasonal variation and correlation with other measures of nursing home quality (correlation with other quality measures is strong concept; seasonal variation is less strong)
Reviewer 6: No concerns
Reviewer 8: Critical Data Elements: Conducted a national validation of MDS 3.0 that tested the criterion validity of the items by examining the agreement between gold-standard nurse assessments and facility nurse assessments based on Kappa statistics.
Performance Measure Score was tested using 1. Correlation with other measures of nursing facility quality including facility CMS five-star rating, health inspections rating, and staffing levels (overall and forRNs); and 2. Seasonal variation.
Reviewer 9: Data element methodology was appropriate. Measure score methodology was weak, but potentially appropriate. Correlating a specific score (weight loss) with a global measure (star rating) is probably not appropriate. Correlating two specific scores (e.g., weight loss to initial patient weight or BMI) at the facility level would be more appropriate.
Reviewer 10: Convergent validity
Reviewer 11: Empirical testing performed.
Reviewer 12: The RAND validation of MDS 3.0 study tested the criterion validity of the items by comparing how different nurses assessed the same residents using MDS 3.0. They compared gold-standard research nurses to gold-standard nurses, and they compare gold-standard nurses to staff nurses trained by the gold-standard nurses. Kappa statistic was calculated.

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

Submission document: Questions 2b.03-04
For example: Is the test sample adequate to generalize for widespread implementation? Do the results demonstrate sufficient validity so that conclusions about quality can be made? Do you agree that the score from this measure as specified is an indicator of quality?
Reviewer 1: See above comments
Reviewer 3: The results of data element testing were excellent with the kappa in the 0.90s.Measure score validity testing showed that correlations with other related quality measures are all in the hypothesized direction.
Reviewer 4: Methods are valid.
Reviewer 5: Data element: kappas ranged between 0.92-0.96, which is very good Measure score: weak correlation coefficients (abs value 0.03 to 0.09) with other measures of quality
Reviewer 6: No concerns
Reviewer 8: Results show the measure has seasonal variation, with highest weigh loss in Q1 and progressively lower rates in Q2-Q4. The correlation results show VERY WEAK negative correlations between the facility-level weight loss QM score and the overall quality rating ($\rho = -.091$, $p < .0001$), health inspection rating ($\rho = -.056$, $p < 0.001$), overall staffing level ($\rho = -.041$, $p < 0.0001$), and RN staffing ($\rho = -0.031$, $p=0.0001$). Although we often see low correlations, these are lower than what we typically see indicating that overall nursing home quality and staffing have little impact on residents likelihood of losing weight. This may indicate the weight loss is more due to patient conditions that nursing home has less control over and not the quality of care provided.
Reviewer 9: The data element validity results were very good. The demonstration that the measure score value had seasonal variation was interesting, but was not persuasive regarding its vailidity. Similarly, the correlation values between the measure score and the global measure (star rating) were both low and negative. This provides further evidence that the comparison using correlation is inappropriate.
Reviewer 10: Stable over time; correlations of convergent validity
Reviewer 11: Adequate sample size.
Reviewer 12: No concerns

VALIDITY: ASSESSMENT OF THREATS TO VALIDITY

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

Submission document: Questions 2b.15-18.

For example: Are there exclusions? If so, are the exclusions/exceptions of sufficient frequency and variation across providers to be needed (and outweigh the data collection burden)? Are any patients or patient groups inappropriately excluded from the measure? If patient preference (e.g., informed decision-making) is a basis for exclusion, does it impact performance and if yes, is the measure specified so that the information about patient preference and the effect on the measure is transparent? If you have concerns based on a clinical rationale, please note here as well as in question #29.

Reviewer 1: None
Reviewer 3: no concerns
Reviewer 4: No concerns
Reviewer 5: none
Reviewer 6: No concerns
Reviewer 9: Measure exclusions (typically end-of-life, hospice) are both conceptually and empirically valid.
Reviewer 10: None
Reviewer 12: No concerns

19. **Risk Adjustment**

Submission Document: Questions 2b.19-32

Applies to all outcome, cost, and resource use measures. Please answer all checkbox questions (19a -19d), then elaborate on your answers in your response to 19e.

19a. **Risk-adjustment method**

☒ None ☐ Statistical model ☐ Stratification
☐ Other method assessing risk factors (please specify)

19b. **If not risk-adjusted, is this supported by either a conceptual rationale or empirical analyses?**

☒ Yes ☒ No ☐ Not applicable

19c. **Social risk adjustment:**

19c.1 Are social risk factors included in risk model? ☐ Yes ☐ No ☒ Not applicable
19c.2 Conceptual rationale for social risk factors included? ☒ Yes ☒ No
19c.3 Is there a conceptual relationship between potential social risk factor variables and the measure focus? ☒ Yes ☒ No

19d.**Risk adjustment summary:**

19d.1 All of the risk-adjustment variables present at the start of care? ☒ Yes ☒ No
19d.2 If factors not present at the start of care, do you agree with the rationale provided for inclusion? ☐ Yes ☒ No
19d.3 Is the risk adjustment approach appropriately developed and assessed? ☒ Yes ☒ No
19d.4 Do analyses indicate acceptable results (e.g., acceptable discrimination and calibration) ☒ Yes ☒ No
19d.5.Appropriate risk-adjustment strategy included in the measure? ☐ Yes ☒ No

19e**. Assess the risk-adjustment approach**

For example: If measure is risk adjusted*:*

- *If the developer asserts there is no conceptual basis for adjusting this measure for social risk factors, do you agree with the rationale?*
- *How well do social risk factor variables that were available and analyzed align with the conceptual description provided?*
- *Are the candidate and final variables included in the risk adjustment model adequately described for the measure to be implemented?*
- *Are all of the risk adjustment variables present at the start of care (if not, do you agree with the rationale)?*
- *If social risk factors are not included in the risk-adjustment approach, do you agree with the developer's decision?*

- *Is an appropriate risk-adjustment strategy included in the measure (e.g., adequate model discrimination and calibration)?*
- *Are all statistical model specifications included, including a "clinical model only" if social risk factors are included in the final model?*

***If measure is NOT risk-adjusted**:*

- *Is a justification for not risk adjusting provided (conceptual and/or empirical)?*
- *Is there any evidence that contradicts the developer's rationale and analysis for not risk-adjusting?*

Reviewer 1: I'm unclear why the measure isn't risk adjusted.
Reviewer 4: The measure developer did not adequately explore the need for risk adjustment. They did not include any comorbidities. Age was specified as a linear variable. The association between weight loss and age is likely to be non-linear, and this should be explored. Patients with certain comorbidities (e.g. cancer) may be more likely to experience weight loss.
Reviewer 5: Measure developer indicated that there was an attempt to develop a risk-adjustment model, but the model explained virtually none of the variance in the weight loss measure.
Reviewer 6: No concerns
Reviewer 8: The measure is not risk adjusted. The stated reason was that the developer's attempt to develop a risk adjusted model were unsuccessful, resulting in low R-Squared. This may reflect the tight range of scores on this measure observed above. This leads to questions as to relevance of this measure. If there are not specific risk factors that may lead to weight loss and could be addressed through appropriate interventions is this a good quality measure? The instability and small variance in rates across facilities suggest it may not be a good measure. That said, a quick lit review indicates there ARE potentially addressable risk factors for unintentional weight loss in long term care facility residents, such as depression, cancer, Parkinson's disease, cognitive impairment, cardiac disorders and benign gastro diseases, eating dependencies, leaves 25% or more of food uneaten, and swallowing/chewing problems (all MDS scored items). It would have been good to see what covariates were tested in the risk adjustment model that had no predictive power at all. It is very surprising that none of these factors were associated with weight loss?
Reviewer 9: Justification was that an attempt to create a risk adjustment model was unsuccessful. NOTE: the submission cites an effort that was conducted > 10 years ago. Another effort using more recent data (e.g., >2014) would be appropriate. Perhaps a stratification approach or some other way that recognizes that weight loss is an important predictor of patient QOL. However, the very small variation among facilities suggests that this measure score may be inappropriate for facility-to-facility comparison."
Reviewer 11: Justification for not risk adjusting provided.
Reviewer 12: Good justification

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

Submission document: Questions 2b.05-07
For cost/resource use measures, does this measure identify meaningful differences about cost and resource use between the measured entities?
Reviewer 3: The measure can identify substantial numbers of performers on both ends.
Reviewer 5: None. 35-40% of facilities have measure performance that is different from national mean.
Reviewer 6: No concerns
Reviewer 8: Due to the low variance in scores across facilities noted above, the instability of the measure, the lack of association with related measures, the lack of risk factors impacting the outcome, I do have concerns about the ability of this measure to identify meaningful differences in performance across facilities as it is currently structured. Weight loss is an important measure and can lead to poor outcomes but this measure does not seem to capture differences in performance in preventing weight loss.
Reviewer 9: There appears to be minimal absolute value differences among facilities in performance on the weight loss measure score.

Reviewer 10: None
Reviewer 11: No concerns.

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

Submission document: Questions 2b.11-14.
***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**. 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. Note if not applicable. Note if applicable but not addressed. If multiple sets of specification (e.g., due to different data sources or methods of data collection): Do analyses indicate they produce comparable results?*
Reviewer 3: no concern
Reviewer 5: Not applicable.
Reviewer 10: None
Reviewer 11: Not applicable

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

Submission document: Questions 2b.08-10.
For example: Are there any sources of missing data not considered? Is it clear how missing data are handled? Is missing data more of a problem for some providers or patients than others? Does the extent of missing data impact the validity of the measure?
Reviewer 5: None. At least 75% of facilities have no missing data on this measure and the mean missing data rate is 1.0%. There is no statistically significant correlation between missing data and QM scores for this measure (r = -.014, p = .12).
Reviewer 6: No concerns
Reviewer 9: Minimal missing data.
Reviewer 10: None
Reviewer 11: No concerns.

For cost/resource use measures ONLY:

If not cost/resource use measure, please skip to question 25.

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

Consider these specific aspects of the measure specifications: attribution, cost categories, target population.

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

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

Attribution: Does the accountable entity have reasonable control over the costs/resources measured? Is this approach aspirational (intending to drive change) or was it developed based on current state?
Costing Approach: Do the cost categories selected align with the measure intent, target population and care settings? Is the approach for assigning dollars to resources
Carve Outs: Has the developer addressed how carve outs in the data source are handled (or should be handled for other users)? For example, if pharmacy data is carved out (missing) from the data set, can a measure that focuses on cost of care for asthmatics still be valid?
Truncation (approach to outliers): What is the threshold for outliers (i.e., extremely high cost or low cost cases) and how are they handled?

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 accountable-entity level testing has been conducted)

☒ **Moderate** (NOTE: Moderate is the highest eligible rating if accountable-entity 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 accountable-entity level and the patient/encounter 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.**

Reviewer 1: The empirical validity testing was adequate. I would have preferred a test of conceptual relationships between facility process, structure, and performance.

Reviewer 3: Empirical measure score validity testing results are in general supportive. This measure is related to other quality measures in the hypothesized way.

Reviewer 4: The measure developer did not adequately explore the need for risk adjustment. They did not include any comorbidities. Age was specified as a linear variable. The association between weight loss and age is likely to be non-linear, and this should be explored. Patients with certain comorbidities (e.g. cancer) may be more likely to experience weight loss.

Reviewer 5: Data element validity testing showed strong results (kappas in the mid 90s). Score level validity testing showed weak results.

Reviewer 7: The results presented demonstrated low correlations and the methods used were maturity level 1

Reviewer 8: See all comments above regarding testing results for validity.

Reviewer 9: Given the minimal differences among facilities, the lack of any risk adjustment to the measure, and the lack of valid discrimination among facilities, this measure lacks validity. Perhaps rethinking the relationship between patient weight loss (vs. BMI or some other measure) and other important nursing home outcomes (e.g., pain medication use; functional behavior) would be more productive than continuing to use this measure.

Reviewer 10: Appear to compare measure to prior versions of itself

Reviewer 11: Sufficient testing completed.

Reviewer 12: No concerns

For composite measures ONLY

If not composite, please skip this section.

Submission documents: Questions 2c.01-08

Examples of analyses:

1) If components are correlated - analyses based on shared variance (e.g., factor analysis, Cronbach's alpha, item-total correlation, mean inter-item correlation).

2) If components are not correlated - analyses demonstrating the contribution of each component to the composite score (e.g., change in a reliability statistic such as ICC, with and without the component measure; change in validity analyses with and without the component measure; magnitude of regression coefficient in multiple regression with composite score as dependent variable, or clinical justification (e.g., correlation of the individual component measures to a common outcome measure).

3) Ideally, sensitivity analyses of the effect of various considered aggregation and weighting rules and the rationale for the selected rules; at a minimum, a discussion of the pros and cons of the considered approaches and rationale for the selected rules.

4) Overall frequency of missing data and distribution across providers. Ideally, sensitivity analysis of the effect of various rules for handling missing data and the rationale for the selected rules; at a minimum, a discussion of the pros and cons of the considered approaches and rationale for the selected rules.

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?**

For example: Do the component measures fit the quality construct and add value? Are the objectives of parsimony and simplicity achieved while supporting the quality construct? Do analyses demonstrate the aggregation and weighting rules fit the quality construct and rationale?

☐ **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.**

1. 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

1ma.01. Indicate whether there is new evidence about the measure since the most recent maintenance evaluation. If yes, please briefly summarize the new evidence, and ensure you have updated entries in the Evidence section as needed.

[Response Begins]
Yes
More recent literature (2015 - 2021) has been included in Section 1a.01 as evidence of the association between unintended weight loss and higher risk of hospitalization and increased mortality. Evidence described in the most recent maintenance evaluation that was not mentioned in past maintenance cycles includes dementia and COVID-19 as risk factors for unintentional weight loss, as well as the role of non-nursing staff in a facility's prevalence of unintentional weight loss. For more details on new evidence about the weight loss measure, please see the red text in Section 1a.01.
[Response Ends]

For more details on new evidence about the weight loss measure, please see the red text between the [Response begins] and [Response Ends] notifications in Section 1a

1a. Evidence

1a.01. Provide a logic model.

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.

[Response Begins]
Outcomes and Risk Factors

[2021 submission] This outcome-based quality measure assesses the percentage of long-stay nursing home residents who experience a weight loss of 5% or more of the baseline weight in the last 30 days, or 10% or more of the baseline weight in the last 6 months, which is not a result of a physician-prescribed weight-loss regimen. Unintended and excessive weight loss is important to monitor in the nursing home population because of the impact on health outcomes, as weight loss is associated with higher risk of hospitalization and increased mortality (Xu et al, 2019; *[2015 submission] Keller et al., 2014; Stack et al., 2013;* [2021 submission] Söderström et al., 2017; Wirth et al., 2016; Wirth et al., 2018). Furthermore, evidence suggests that weight loss is associated with increased mortality in the elderly population regardless of baseline BMI (Pizzato et al., 2015).

Some nursing home residents are at higher risk for experiencing weight loss, as certain resident characteristics influence the rate of weight loss. Nursing home residents who have a low BMI, are older in age (85-94), or have trouble with meal intake are at high risk for experiencing weight loss (Torbahn et al, 2021; Sanford et al., 2020). Researchers have further explored difficulties with meal intake among the nursing home population and found that residents who have eating dependencies, are prescribed pureed or liquidized diets, or experience difficulties swallowing due to conditions such as dysphagia are at higher risk for experiencing weight loss than their counterparts (Beck, A. M., 2015; Keller et al., 2017; Wirth et al., 2018). In fact, one cross-sectional study observed significantly lower BMIs and a significantly higher six month mortality rate for nursing home residents with dysphagia than those without dysphagia (Wirth et al., 2018). Another

contributing factor to difficulties with meal intake is a resident's masticatory function. Studies have shown that the absence of teeth or dentures negatively impacts residents' masticatory function and is one of the most influential factors on nutritional risk (de Medeiros et al., 2020; Schmalz et al., 2021). In fact, a 2021 study by de Medeiros et al. observed that edentulous nursing home residents who did not have complete dentures demonstrated worse masticatory function than edentulous residents with complete dentures, partially dentate residents with prosthesis, and partially dentate residents without prosthesis. The researchers concluded that any absence of teeth and dentures negatively impacts both a resident's swallowing threshold and masticatory function, thereby increasing their risk for malnutrition and weight loss (de Medeiros et al., 2020). Additionally, residents who have depression, severe cognitive impairment, or low functional status are at high risk for experiencing weight loss (Velázquez-Alva et al., 2020; Madeira et al., 2019). A 2020 study by Velázquez-Alva found that residents with depression were five times more likely to have malnutrition or be at risk for malnutrition while residents with better nutritional status had lower depression scores. This evidence suggests that residents with depression may lose interest in or refuse to eat, which could lead to unintended weight loss.

Another potential risk factor for weight loss in the nursing home population is a diagnosis of dementia. A 2017 prospective cohort study by de Souto Barreto et al. examined the obesity paradox, which they defined as the phenomenon where "the risk of all-cause death in subjects aged 65 years or over is observed to be lower in those with a high BMI and higher in those with a low BMI" when compared to people with normal BMI, among people living with dementia (PLWD) and people who are not living with dementia in France's nursing home population (2017). The researchers categorized their subjects according to BMI as underweight, normal weight, overweight, and obese and followed them over time. The results of the study demonstrated that mortality risk for PLWD was reduced by almost half for those who were overweight and obese, while the mortality risk for those who were underweight increased when compared to PLWD who had normal weights (de Suoto Barreto et al., 2017). This evidence suggests dementia may amplify the effects of the obesity paradox in the elderly population, and therefore it is important for facilities to strive to prevent weight loss for any PLWD in their care.

A final potential risk factor for unintended weight loss among nursing home residents is the COVID-19 virus. A 2021 retrospective cohort study compared nursing home long-stay resident health outcomes observed between March and July 2020 to outcomes observed between 2017 and 2019 to quantify the effects of the COVID-19 pandemic (Levere, Rowan, & Wysocki). Unintended substantial weight loss in the long-stay nursing home population increased by 6 percentage points during the pandemic (Levere, Rowan, & Wysocki, 2021). Although this unintended weight loss was greatest among residents who contracted COVID-19 (roughly 10% of residents each week), during each week of the pandemic study window roughly 7.5% of residents who did not contract COVID-19 also experienced unintended weight loss (Levere, Rowan, & Wysocki, 2021). Similar findings were observed by a single skilled nursing facility (SNF) in Chicago, Illinois where residents who contracted COVID-19 experienced weight loss of 4.6% and residents who did not contract COVID-19 experienced weight loss of 2.4% between March 1, 2020 and May 31, 2020 (Martinchek et al., 2020). This study is limited by small sample size, but the results warrant further investigation as the impact of the ongoing pandemic continues to be assessed. However, the currently available evidence suggests that COVID-19 is a risk factor for weight loss among nursing home residents regardless of their infection status.

Evidence for link between structure and quality of care outcomes

Several nursing home characteristics may influence the risk for experiencing weight loss, including adequate staffing levels and staff training and education. *[2015 submission] To help prevent unintended weight loss,* [2021 submission] it is essential that staff dedicate sufficient time to assisting and feeding residents *[2015 submission] who cannot feed themselves, [2021 submission]* as facilities with three or more hours of nurse's aide (NA) time per resident per day have decreased rates of unintended weight loss (Tuinman et al., 2021). *[2015 submission] Higher staffing levels and staff training, positive relationships and better communication between staff and residents, and a quieter and more private dining environment may increase residents' food consumption and improve their nutritional status (Beattie et al., 2014; Pelletier, 2004; Altus et al., 2002; Simmons et al., 2001; Amella, 1999; Kayser-Jones et al., 1997; Van Ort et al., 1995; Lange-Alberts et al., 1994; Sanders, 1990).* [2021 submission] Additionally, staff turnover and tenure influence the prevalence of weight loss in a facility. A 2019 study by Juh Hyun Shin found that tenure of more than one year for registered nurses (RNs) contributed to a lower prevalence of weight loss among a facility's residents, whereas high turnover rates for social workers aligned with increased rates of weight loss. Certified nursing assistants (CNAs) are

essential staff that also play a pivotal role in determining the prevalence of weight loss in a facility. A 2017 cross-sectional study by Trinkoff et al. comparing CNA training requirements and nursing home resident outcomes observed a statistically significant association between higher in-service general training requirements and lower rates of weight loss in small facilities.

Non-nursing staff can also play a key part in determining the prevalence of weight loss in a facility. Recent studies have shown that if trained thoroughly in feeding assistance, non-nursing staff can have a positive and substantial effect on increasing caloric intake among nursing home residents (Simmons et al., 2017; Hollingsworth, Long, & Simmons, 2018). In a recent randomized controlled trial, researchers compared the quality of feeding assistance provided by trained non-nursing staff with care provided by CNAs and found that trained non-nursing staff performed significantly better than CNAs for 12 of 13 care process measures, all of which aligned with common dietary and dignity federal regulations known as F-tags (Hollingsworth, Long, & Simmons, 2018). Therefore, interventions aimed at providing quality feeding assistance trainings to all staff may considerably reduce the prevalence of unintended weight loss in facilities.

Another *[2015 submission] nursing home characteristic* [2021 submission] to consider is the physical environment of a facility, which has *[2015 submission] been shown to be associated with the nutritional status and weight loss of residents. Communication between staff and residents, verbal prompting, positive reinforcement, and adequate time for meals all contribute to increased food consumption (Pelletier, 2004; Altus et al., 2002; Simmons, Osterweil et al., 2001; Simmons et al., 2001; Amella et al., 1999; Kayser-Jones et al., 1997; Van Ort et al., 1995; Lange-Alberts et al., 1994; Sanders, 1990). In fact, one study found facilities that utilize verbal prompting and social interaction during meals had a lower rate of weight loss (Simmons et al., 2003). Additionally, noisy, chaotic, and institutional dining rooms are associated with low consumption of food and drink (Reed et al., 2005; Durnbaugh et al., 1996; Van Ort et al., 1995).*

Evidence for link between processes and quality of care outcomes

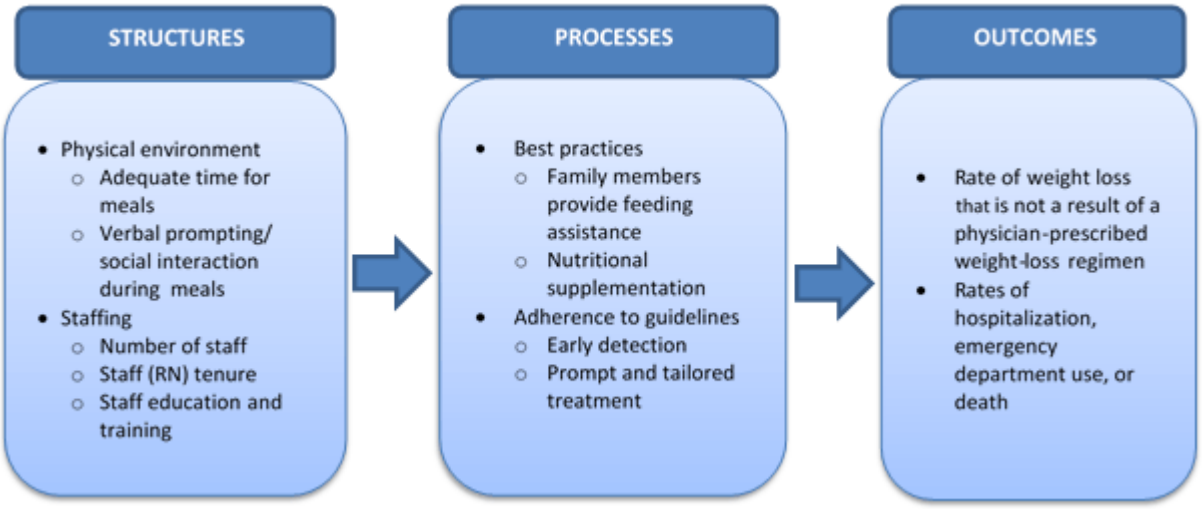
[2021 submission] Key nursing home processes have been found to influence the rate of weight loss within a facility. These key processes include adherence to clinical guidelines and best practices in weight loss treatment. Clinical practice guidelines for treating weight loss emphasize the need for early detection and treatment, as failure to promptly identify unintended weight loss and its causes can have serious implications on morbidity and mortality. Therefore, the Society for Post-Acute and Long-Term Care Medicine (AMDA) recommends evaluating a resident's nutritional status as soon as possible after admission, which includes recording weight, height, BMI, eating preferences, and other baseline testing (2020). After this initial evaluation, facility staff should observe residents routinely for changes in weight or food intake in order to accurately diagnose and treat weight loss according to AMDA's criteria and recommended interventions. Moreover, AMDA argues that the interventions outlined in their clinical practice guideline are intended to be customized to meet the needs of individual residents (2020). Customization of these guidelines translates to individualizing meal plans to meet residents' personal preferences, as well as their ability to feed themselves (AMDA, 2020). AMDA also suggests that facilities implement feeding programs that involve a variety of caregivers, such as facility staff, family members, and volunteers to promote increased caloric intake (2020).

AMDA's recommendation to implement feeding programs that include family members is further supported by a 2020 cross-sectional study that found family assistance with feeding was associated with significantly higher consumption of protein and calories (Wu et al.). Researchers observed residents consuming significantly more calories (50kcal/day) and protein (2.7g/day) when family members provided eating assistance than when facility staff provided eating assistance to the same residents (Wu et al., 2020). This increase of 50 calories per day would contribute to an approximate increase in body weight of five pounds over the course of one year, which is clinically meaningful and can reduce the risk of malnutrition (Wu et al., 2020).

In addition to feeding programs that involve family members providing feeding assistance to residents, other *[2015 submission] nutrition and dining programs may potentially reduce the risk of weight loss for nursing home residents. For example, a Cochrane meta-analysis found that supplementation produces small but consistent weight gain in older people (Milne et al., 2005).* [2021 submission] Additionally, a more recent study that assessed the impact of oral nutritional supplements on nutritional and functional status in malnourished nursing home residents found that after receiving two bottles of an energy-rich, high-protein commercial supplement daily for three months, residents experienced significant

increases in body weight and BMI and lower risk for malnutrition (Malafarina et al., 2021). Although residents who are experiencing weight loss or malnutrition would benefit from nutritional supplementation, the likelihood of receiving such supplements may be impacted by certain facility characteristics. Facilities that have a dietician on staff, have a coordinating physician with specific training in geriatrics on staff, and organize an evening snack for residents are found to be associated with the prescription of supplements (Dupuy et al., 2016).

Figure 1: Role of Nursing Home Structures and Processes in Rates of Weight Loss



References:

[2015 submission] Altus, D. E., Engelman, K. K., & Mathews, R. M. (2002). Using family-style meals to increase participation and communication in persons with dementia. Journal of Gerontological Nursing, 28(9), 47-53.

Amella, E. J. (1999). Factors influencing the proportion of food consumed by nursing home residents with dementia. Journal of the American Geriatrics Society, 47(7), 879-885.

Beattie, E., O'Reilly, M., Strange, E., Franklin, S., & Isenring, E. (2014). How much do residential aged care staff members know about the nutritional needs of residents? Int J Older People Nurs, 9(1), 54-64. doi: 10.1111/opn.12016

[2021 submission] Beck, A.M. (2015). Weight loss, mortality and associated potentially modifiable nutritional risk factors among nursing home residents — A Danish follow-up study. Journal of Nutrition, Health and Aging; 96–101. https://doi.org/10.1007/s12603-015-0439-6

de Medeiros MMD, Pinheiro MA, de Figueredo OMC, de Oliveira LFS, Wanderley RL, Cavalcanti YW, Rodrigues Garcia RCM. Masticatory function in nursing home residents: Correlation with the nutritional status and oral health-related quality of life. Journal of Oral Rehabilitation. 2020 Dec;47(12):1511-1520. doi: 10.1111/joor.13096. Epub 2020 Oct 5. PMID: 32964471.

de Souto Barreto, P., Cadroy, Y., Kelaiditi, E., Vellas, B., Rolland, Y. (2017). The prognostic value of body-mass index on mortality in older adults with dementia living in nursing homes. Clinical Nutrition; 36(2), 423-428. doi: 10.1016/j.clnu.2015.12.009. Epub 2015 Dec 18. PMID: 26724185.

Dupuy, C., de Souto Barreto, P., Ghisolfi, A., Guyonnet, S., Dorigny, B., Vellas, B., Rolland, Y. (2016). Indicators of oral nutritional supplements prescription in nursing home residents: A cross-sectional study. Clinical Nutrition; 35(5):1047-52. doi: 10.1016/j.clnu.2015.07.015. Epub 2015 Jul 28. PMID: 26243061.

[2015 submission] Durnbaugh, T., Haley, B., & Roberts, S. (1996). Assessing problem feeding behaviors in mid-stage Alzheimer's disease: Clients with mid-stage Alzheimer's disease may be eating far less than their caregivers believe. Geriatric Nursing, 17(2), 63-67.

[2021 submission] Hollingsworth, E.K., Long, E.A., Simmons, S.F. (2018). Comparison Between Quality of Care Provided by Trained Feeding Assistants and Certified Nursing Assistants During Between-Meal Supplementation in Long-Term Care Settings. Journal of Applied Gerontology; 37(11):1391-1410. doi: 10.1177/0733464816669806. Epub 2016 Sep 22. PMID: 27664171.

[2015 submission] Kayser-Jones, J., & Schell, E. (1997). The mealtime experience of a cognitively impaired elder: ineffective and effective strategies. Journal of Gerontological Nursing, 23(7), 33-39.

Keller, H., Beck, A. M., & Namasivayam, A. (2014). Improving food and fluid Intake for older adults living in long-term care: A research agenda. Journal of the American Medical Directors Association.
[2021 submission] Keller, H.H., Carrier, N., Slaughter, S.E., Lengyel, C., Steele, C.M., Duizer, L., Morrison, J., Brown, K.S., Chaudhury, H., Yoon, M.N., Duncan, A.M., Boscart, V., Heckman, G., Villalon, L. (2017). Prevalence and Determinants of Poor Food Intake of Residents Living in Long-Term Care. Journal of the American Medical Directors Association; 18(11):941-947. doi: 10.1016/j.jamda.2017.05.003. Epub 2017 Jun 28. PMID: 28668663.
[2015 submission] Lange-Alberts, M. E., & Shott, S. (1994). Nutritional intake. Use of touch and verbal cuing. Journal of Gerontological Nursing, 20(2), 36-40.
[2021 submission] Levere, M., Rowan, P., Wysocki, A. (2021). The Adverse Effects of the COVID-19 Pandemic on Nursing Home Resident Well-Being. Journal of the American Medical Directors Association; 22(5):948-954.e2. doi: 10.1016/j.jamda.2021.03.010. Epub 2021 Mar 20. PMID: 33861980; PMCID: PMC7980137.
Madeira, T., Peixoto-Plácido, C., Sousa-Santos, N., Santos, O., Alarcão, V., Goulão, B., . . . Gorjão Clara, J. (2019). Malnutrition among older adults living in Portuguese nursing homes: The PEN-3S study. Public Health Nutrition, 22(3), 486-497. doi:10.1017/S1368980018002318
Malafarina, V., Serra Rexach, J.A., Masanés, F., Cervera-Díaz, M.C., Lample Lacasa, L., Ollero Ortigas, A., Cruz-Jentoft, A.J. (2021). Results of High-Protein, High-Calorie Oral Nutritional Supplementation in Malnourished Older People in Nursing Homes: An Observational, Multicenter, Prospective, Pragmatic Study (PROT-e-GER). Journal of the American Medical Directors Association; S1525-8610(21)00297-8. doi: 10.1016/j.jamda.2021.02.039. Epub ahead of print. PMID: 33819452.
[2015 submission] Martinchek, M., Beiting, K.J., Walker, J., Thompson, K., Gleason, L.J., Levine, S. (2020). Weight loss in COVID-19-positive nursing home residents. Journal of the American Medical Directors Association; 22(2):257-258.e1. doi:10.1016/j.jamda.2020.11.031
Milne AC, Potter J, Vivanti A, Avenell A. Protein and energy supplementation in elderly people at risk from malnutrition. Cochrane Database of Systematic Reviews 2009, Issue 2. Art. No.: CD003288. DOI: 10.1002/14651858.CD003288.pub3: http://summaries.cochrane.org/CD003288/ENDOC_protein-and-energy-supplementation-in-elderly-people-at-risk-from-malnutrition
Pelletier, C. A. (2004). What do certified nurse assistants actually know about dysphagia and feeding nursing home residents?. American Journal of Speech-Language Pathology, 13(2), 99-113.
[2021 submission] Pizzato, S., Sergi, G., Bolzetta, F., De Rui, M., De Ronch, I., Carraro, S., Berton, L., Orr, E., Imoscopi, A., Perissinotto, E., Coin, A., Manzato, E., Veronese, N. (2015). Effect of weight loss on mortality in overweight and obese nursing home residents during a 5-year follow-up. European Journal of Clinical Nutrition; 69(10):1113-8. doi: 10.1038/ejcn.2015.19. Epub 2015 Mar 11. PMID: 25758838
[2015 submission] Reed, P. S., Zimmerman, S., Sloane, P. D., Williams, C. S., & Boustani, M. (2005). Characteristics associated with low food and fluid intake in long-term care residents with dementia. The Gerontologist, 45(suppl 1), 74-81.
Sanders, H. (1990). Nutrition management in long-term care. J Nutr Elder, 9(3), 69-74.
[2021 submission] Sanford, A.M., Morley, J.E., Berg-Weger, M., Lundy, J., Little, M.O., Leonard, K., Malmstrom, T.K. (2020). High prevalence of geriatric syndromes in older adults. PLoS ONE 15(6): e0233857. https://doi.org/10.1371/journal.pone.0233857
Schmalz G, Denkler CR, Kottmann T, Rinke S, Ziebolz D. Oral Health-Related Quality of Life, Oral Conditions, and Risk of Malnutrition in Older German People in Need of Care-A Cross-Sectional Study. Journal of Clinical Medicine. 2021 Jan 22;10(3):426. doi: 10.3390/jcm10030426. PMID: 33499288; PMCID: PMC7866095.
Shin, J.H. (2019) Nursing Staff Characteristics on Resident Outcomes in Nursing Homes. Journal of Nursing Research. 27(1):1-9. doi:10.1097/jnr.0000000000000269
[2015 submission] Simmons, S. F., Alessi, C., & Schnelle, J. F. (2001). An intervention to increase fluid intake in nursing home residents: prompting and preference compliance. Journal of the American Geriatrics Society, 49(7), 926-933.
Simmons, S. F., Garcia, E. T., Cadogan, M. P., Al-Samarrai, N. R., Levy-Storms, L. F., Osterweil, D., & Schnelle, J. F. (2003). The Minimum Data Set Weight-Loss Quality Indicator: Does It Reflect Differences in Care Processes Related to Weight Loss?. Journal of the American Geriatrics Society, 51(10), 1410-1418.
[2021 submission] Simmons, S.F., Hollingsworth, E.K., Long, E.A., Liu, X., Shotwell, M.S., Keeler, E., An, R., Silver, H,J. (2017). Training Nonnursing Staff to Assist with Nutritional Care Delivery in Nursing Homes: A Cost-Effectiveness Analysis. Journal of the American Geriatrics Society; 65(2):313-322. doi: 10.1111/jgs.14488. Epub 2016 Nov 7. PMID: 28198565; PMCID: PMC5315268.
Söderström, L., Rosenblad, A., Thors Adolfsson, E., & Bergkvist, L. (2017). Malnutrition is associated with increased mortality in older adults regardless of the cause of death. British Journal of Nutrition,117(4), 532-540. doi:10.1017/S0007114517000435

[2015 submission] Stack, S., Chertow, G. M., Johansen, K. L., Si, Y., & Tamura, M. K. (2013). Pre-ESRD changes in body weight and survival in nursing home residents starting dialysis. Clinical Journal of the American Society of Nephrology, 8(10), 1734-1740.
[2021 submission] Torbahn, G., Sulz, I., Großhauser, F., Hiesmayr, M.J., Kiesswetter, E., Schindler, K., Sieber, C.C., Visser, M., Weber, J., Volkert, D. (2021). Predictors of incident malnutrition-a nutritionDay analysis in 11,923 nursing home residents. European Journal of Clinical Nutrition. doi: 10.1038/s41430-021-00964-9. Epub ahead of print. PMID: 34239065.
Trinkoff, A.M., Storr, C.L., Lerner, N.B., Yang, B.K., Han, K. (2017). CNA Training Requirements and Resident Care Outcomes in Nursing Homes. The Gerontologist; 57(3):501-508. doi: 10.1093/geront/gnw049. PMID: 27059825.
Tuinman, A., De Greef, M.H.G., Finnema, E.J., Roodbol, P.F. (2021). A systematic review of the association between nursing staff and nursing-sensitive outcomes in long-term institutional care. Journal of Advanced Nursing; 77(8):3303-3316. doi: 10.1111/jan.14840. Epub 2021 Mar 25. PMID: 33764569.
[2015 submission] Van Ort, S., & Phillips, L. R. (1995). Nursing intervention to promote functional feeding. Journal of Gerontological Nursing, 21(10), 6-14.
[2021 submission] Velázquez-Alva, M.C., Irigoyen-Camacho, M.E., Cabrer-Rosales, M.F., Lazarevich, I., Arrieta-Cruz, I., Gutiérrez-Juárez, R., Zepeda-Zepeda, M.A. (2020). Prevalence of Malnutrition and Depression in Older Adults Living in Nursing Homes in Mexico City. Nutrients; 12(8), 2429. https://doi.org/10.3390/nu12082429
Wirth, R., Streicher, M., Smoliner, C., Kolb, C., Hiesmayr, M., Thiem, U., Sieber, CC., Volkert, D. (2016). The impact of weight loss and low BMI on mortality of nursing home residents - Results from the nutritionDay in nursing homes. Clinical Nutrition: 35(4), 900-6. doi: 10.1016/j.clnu.2015.06.003. Epub 2015 Jun 19. PMID: 26143743.
Wirth, R., Pourhassan, M., Streicher, M., Hiesmayr, M., Schindler, K., Sieber, C.C., Volkert, D. (2018). The Impact of Dysphagia on Mortality of Nursing Home Residents: Results From the nutritionDay Project. Journal of the American Medical Directors Association; 19(9):775-778. doi: 10.1016/j.jamda.2018.03.016. Epub 2018 May 31. PMID: 29778638.
Wu, S.A., Morrison-Koechl, J., Slaughter, S.E., Middleton, L.E., Carrier, N., McAiney, C., Lengyel, C., Keller, H. (2020). Family member eating assistance and food intake in long-term care: A secondary data analysis of the M3 Study. Journal of Advanced Nursing; 76(11):2933-2944. doi: 10.1111/jan.14480. Epub 2020 Sep 4. PMID: 32885494.
Xu, D., Kane, R., Arling, G. (2019). Relationship between nursing home quality indicators and potentially preventable hospitalisation. BMJ Quality & Safety; 28(7):524-533. doi: 10.1136/bmjqs-2018-008924. Epub 2019 Mar 13. PMID: 30867234.

[Response Ends]

1a.03. Provide empirical data demonstrating the relationship between the outcome (or PRO) and at least one healthcare structure, process, intervention, or service.

[Response Begins]
[2021 submission] Facility staffing is one health care structural indicator that plays an important role in the rates of weight loss in the nursing home population. The Payroll Based Journal (PBJ) Nursing Staffing and Non-Nursing Staffing contains publicly reported data on the number of hours staff are paid to work each day by facility. Staffing levels are also evaluated, rated, and publicly reported as part of CMS's Five-Star Quality Rating System. This system was designed to help consumers, their families, and caregivers compare nursing homes more easily and to help identify areas about which consumers may have questions. Staffing is one of three sources used by the system to determine a facility's over five-star rating. The correlations between several of these staffing ratings and the weight loss measure were assessed with the hypothesis that staffing ratings would be inversely related to a facility's percentile ranking for the weight loss measure. The results of this convergent validity testing are summarized in Table 1 below, which indicates that NQF 0689 has small but statistically significant negative correlations with staffing ratings and registered nurse staffing ratings. These correlations are in the expected direction and suggest that lower overall staffing ratings are associated with increased rates of weight loss and lower RN staffing ratings are also associated with increased rates of weight loss.

Table 1: Convergent Validity of NQF 0689 and Facility Staffing

Quality Measure	Spearman Correlation	P-Value
Facility five-star ratings	-	-
Staffing ratings	-0.029	<.0001
Registered nurse (RN) staffing ratings	-0.011	<.0001

Cells marked by a dash (-) are intentionally left blank.

[Response Ends]

1b. Performance Gap

1b.01. Briefly explain the rationale for this measure.

Explain how the measure will improve the quality of care, and list the benefits or improvements in quality envisioned by use of this measure.

[Response Begins]

This outcome-based quality measure the percentage of long-stay nursing home residents who experience a weight loss of 5% or more of the baseline weight in the last 30 days, or 10% or more of the baseline weight in the last 6 months, which is not a result of a physician-prescribed weight-loss regimen. Unintended and excessive weight loss is important to monitor in the nursing home population because of the impact on health outcomes, as weight loss is associated with higher risk of hospitalization and increased mortality (Xu et al, 2019; Keller et al., 2014; Stack et al., 2013; Söderström et al., 2017; Wirth et al., 2016; Wirth et al., 2018). Furthermore, evidence suggests that weight loss is associated with increased mortality in the elderly population regardless of baseline BMI (Pizzato et al., 2015). Additionally, studies have shown that risk for unintended weight loss is associated with a variety of resident characteristics, including but not limited to, increasing age, low BMI, eating dependency, absence of teeth or dentures, depression, severe cognitive impairment, low functional status, and dementia (Torbahn et al, 2021; Sanford et al., 2020; Beck, A. M., 2015; Keller et al., 2017; Wirth et al., 2018; Velázquez-Alva et al., 2020; Madeira et al., 2019; de Souto Barreto et al., 2017). The capacity of nursing homes to provide residents sufficient support, such as feeding assistance and nutritional supplements, by appropriately trained nursing and non-nursing staff to serve the medical needs of residents can mitigate nutritional risks and prevent unintended weight loss.

References:

Beck, A.M. (2015). Weight loss, mortality and associated potentially modifiable nutritional risk factors among nursing home residents — A Danish follow-up study. Journal of Nutrition, Health and Aging; 96–101. https://doi.org/10.1007/s12603-015-0439-6

de Souto Barreto, P., Cadroy, Y., Kelaiditi, E., Vellas, B., Rolland, Y. (2017). The prognostic value of body-mass index on mortality in older adults with dementia living in nursing homes. Clinical Nutrition; 36(2), 423-428. doi: 10.1016/j.clnu.2015.12.009. Epub 2015 Dec 18. PMID: 26724185.

Keller, H., Beck, A. M., & Namasivayam, A. (2014). Improving food and fluid Intake for older adults living in long-term care: A research agenda. Journal of the American Medical Directors Association.

Madeira, T., Peixoto-Plácido, C., Sousa-Santos, N., Santos, O., Alarcão, V., Goulão, B., . . . Gorjão Clara, J. (2019). Malnutrition among older adults living in Portuguese nursing homes: The PEN-3S study. Public Health Nutrition, 22(3), 486-497. doi:10.1017/S1368980018002318

Pizzato, S., Sergi, G., Bolzetta, F., De Rui, M., De Ronch, I., Carraro, S., Berton, L., Orr, E., Imoscopi, A., Perissinotto, E., Coin, A., Manzato, E., Veronese, N. (2015). Effect of weight loss on mortality in overweight and obese nursing home residents during a 5-year follow-up. European Journal of Clinical Nutrition; 69(10):1113-8. doi: 10.1038/ejcn.2015.19. Epub 2015 Mar 11. PMID: 25758838

Sanford, A.M., Morley, J.E., Berg-Weger, M., Lundy, J., Little, M.O., Leonard, K., Malmstrom, T.K. (2020). High prevalence of geriatric syndromes in older adults. PLoS ONE 15(6): e0233857. https://doi.org/10.1371/journal.pone.0233857

Söderström, L., Rosenblad, A., Thors Adolfsson, E., & Bergkvist, L. (2017). Malnutrition is associated with increased mortality in older adults regardless of the cause of death. British Journal of Nutrition,117(4), 532-540. doi:10.1017/S0007114517000435
Stack, S., Chertow, G. M., Johansen, K. L., Si, Y., & Tamura, M. K. (2013). Pre-ESRD changes in body weight and survival in nursing home residents starting dialysis. Clinical Journal of the American Society of Nephrology, 8(10), 1734-1740.
Torbahn, G., Sulz, I., Großhauser, F., Hiesmayr, M.J., Kiesswetter, E., Schindler, K., Sieber, C.C., Visser, M., Weber, J., Volkert, D. (2021). Predictors of incident malnutrition-a nutritionDay analysis in 11,923 nursing home residents. European Journal of Clinical Nutrition. doi: 10.1038/s41430-021-00964-9. Epub ahead of print. PMID: 34239065.
Velázquez-Alva, M.C., Irigoyen-Camacho, M.E., Cabrer-Rosales, M.F., Lazarevich, I., Arrieta-Cruz, I., Gutiérrez-Juárez, R., Zepeda-Zepeda, M.A. (2020). Prevalence of Malnutrition and Depression in Older Adults Living in Nursing Homes in Mexico City. Nutrients; 12(8), 2429. https://doi.org/10.3390/nu12082429
Wirth, R., Streicher, M., Smoliner, C., Kolb, C., Hiesmayr, M., Thiem, U., Sieber, CC., Volkert, D. (2016). The impact of weight loss and low BMI on mortality of nursing home residents - Results from the nutritionDay in nursing homes. Clinical Nutrition: 35(4), 900-6. doi: 10.1016/j.clnu.2015.06.003. Epub 2015 Jun 19. PMID: 26143743.
Wirth, R., Pourhassan, M., Streicher, M., Hiesmayr, M., Schindler, K., Sieber, C.C., Volkert, D. (2018). The Impact of Dysphagia on Mortality of Nursing Home Residents: Results From the nutritionDay Project. Journal of the American Medical Directors Association; 19(9):775-778. doi: 10.1016/j.jamda.2018.03.016. Epub 2018 May 31. PMID: 29778638.
Xu, D., Kane, R., Arling, G. (2019). Relationship between nursing home quality indicators and potentially preventable hospitalisation. BMJ Quality & Safety; 28(7):524-533. doi: 10.1136/bmjqs-2018-008924. Epub 2019 Mar 13. PMID: 30867234.

[Response Ends]

1b.02. Provide performance scores on the measure as specified (current and over time) at the specified level of analysis.

Include mean, std dev, min, max, interquartile range, and 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 (4b) under Usability and Use.

[Response Begins]
Current performance: Table 8 in the "Scientific Acceptability: Validity - Threats to Validity (Statistically Significant Differences, Multiple Data Sources, Missing Data)" section of this form describes the national facility score distribution for Percent of Residents Who Lose Too Much Weight. The four-quarter facility-level mean score for this measure between Quarter 1 (Q1) and Quarter 4 (Q4) of 2019 was 5.2% and the median score was 4.8%. The standard deviation was 3.1%, the score at the 10th percentile was 1.6%, and the score at the 90th percentile was 9.2%. The interquartile range for this measure was 3.9%, indicating some room for improvement in this measure. Of the facilities with adequate sample size to report, 2.6% had perfect scores of 0. This analysis is restricted to facilities that had at least 20 residents in the denominator, the minimum denominator threshold for public reporting. Between 2019Q1 and 2019Q4, there were 14,274 facilities (93.1%) and 1,230,998 residents (98.6%) that met the denominator inclusion criteria.

n (Facilities): 14,274

k (Residents): 1,230,998

Mean score: 5.2%

Std dev.: 3.1%

10th percentile: 1.6%

25th percentile: 3.1%

50th percentile: 4.8%

75th percentile: 6.9%

90th percentile: 9.2%

Interquartile range: 3.9%.

% of facilities with “perfect scores”: 2.6%

Performance Over Time: The national facility-level mean and median scores for the Percent of Residents Who Lose Too Much Weight demonstrate stability from quarter to quarter (Figure 1 in the “Scientific Acceptability: Validity – Testing” section of this form). Overall, the national facility-level mean and median scores have remained relatively stable with only a slight increase in 2019Q2. The mean score for this measure was 5.24% in 2018Q1 and the median score was 4.55%. In 2019Q2, the mean and median were 5.44% and 4.76%, respectively. (Data Source: Data are drawn from all United States Nursing Homes with Medicare certified beds and a minimum of 20 long-stay residents in their denominator in each quarter.)
[Response Ends]

1b.03. If no or limited performance data on the measure as specified is reported above, 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. Include citations.

[Response Begins]
This is not applicable (data are available and described in 1b.02).
[Response Ends]

1b.04. 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.

Describe the data source including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities included. Include mean, std dev, min, max, interquartile range, and scores by decile. 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 (4b) under Usability and Use.

[Response Begins]
Age

To examine whether facilities with higher percentages of residents aged 85 or older have different performance scores for too much weight loss, analyses were completed comparing the performance of facilities based on their percentage of residents aged 85 or older and residents below the age of 85. First, the percentage of residents who lose too much weight was stratified by age. Residents aged 85 or older represented the highest mean (5.9%), followed by residents below the age of 85 (5.1%). Next, a 2-way chi-squared test for statistical dependence was run that assessed the association between quality measure score and age. The results were significant (p=<.0001) indicating that there is a statistically significant relationship between age and QM score for the measure. The results suggested that residents aged 85 years or older are at higher risk of experiencing too much weight loss than residents less than 85 years of age.

Race

To examine whether facilities with higher percentages of non-white residents have different performance scores for too

much weight loss, analyses were completed comparing the performance of facilities based on their percentage of white only and non-white residents. First, the percentage of residents who lose too much weight was stratified by racial identification. White residents represented the highest mean (5.40%), followed by Black or African American residents (5.38%), Asian residents (5.13%), Hispanic or Latino residents (4.81%), American Indian/Alaska Native residents (4.76%), and Native Hawaiian or Other Pacific Islander residents (4.75%). Next a 2-way chi-squared test for statistical dependence was run that assessed the association between quality measure score and race/ethnicity. The results were significant (p=<.0001) indicating that there is a statistically significant relationship between racial composition and QM score for the measure. The results suggested that the white only population (5.4%) is at slightly higher risk for losing too much weight than the non-white only population (5.2%).

Socioeconomic status

To examine whether facilities with higher percentages of Medicaid-enrolled residents have different performance scores for too much weight loss, analyses were completed comparing the performance of facilities based on their percentage of Medicaid-enrolled residents, including dual beneficiaries who are also enrolled in Medicare, and residents not enrolled in Medicaid. First, the percentage of residents who lose too much weight was stratified by Medicaid enrollment. Residents not enrolled in Medicaid represented the highest mean (6.27%), followed by Medicaid-enrolled residents (5.20%), indicating there are slightly more residents not enrolled in Medicaid who lose too much weight than Medicaid-enrolled residents. Next, a 2-way chi-squared test for statistical dependence was run that assessed the association between quality measure score and Medicaid enrollment. The results were significant (p=<.0001) indicating that that there is a statistically significant relationship between Medicaid enrollment and QM score for this measure. The results suggested that the non-Medicaid population is at higher risk of losing too much weight than the Medicaid population, indicating there is a relationship between socioeconomic status and too much weight loss among long-stay residents.

SOURCE: Acumen analysis of Q4 2019 MDS 3.0 data
[Response Ends]

1b.05. If no or limited data on disparities from the measure as specified is reported above, 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 above.

[Response Begins]
This is not applicable (data are available and described in Section 1b.04).
[Response Ends]

2. 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.

---

sp.01. Provide the measure title.

Measure titles should be concise yet convey who and what is being measured (see What Good Looks Like).

[Response Begins]
Percent of Residents Who Lose Too Much Weight (Long-Stay)
[Response Ends]

sp.02. Provide a brief description of the measure.

Including type of score, measure focus, target population, timeframe, (e.g., Percentage of adult patients aged 18-75 years receiving one or more HbA1c tests per year).

[Response Begins]
This measure captures the percentage of long-stay nursing home residents with a target Minimum Data Set (MDS) 3.0 assessment (OBRA, PPS, or discharge) that indicates a weight loss of 5% or more of the baseline weight in the last 30 days, or 10% or more of the baseline weight in the last 6 months, which is not a result of a physician-prescribed weight-loss regimen. The baseline weight is the resident's weight closest to 30 or 180 days before the date of the target assessment. Long-stay nursing facility residents are identified as those who have had 101 or more cumulative days of nursing facility care.
[Response Ends]

sp.04. Check all the clinical condition/topic areas that apply to your measure, below.

Please refrain from selecting the following answer option(s). We are in the process of phasing out these answer options and request that you instead select one of the other answer options as they apply to your measure.

Please do not select:

- *Surgery: General*

[Response Begins]
Other (specify)
This measure is for non-prescribed weight loss.
[Response Ends]

sp.05. Check all the non-condition specific measure domain areas that apply to your measure, below.

[Response Begins]
Health and Functional Status: Nutrition
[Response Ends]

sp.06. Select one or more target population categories.

Select only those target populations which can be stratified in the reporting of the measure's result.

Please refrain from selecting the following answer option(s). We are in the process of phasing out these answer options and request that you instead select one of the other answer options as they apply to your measure.

Please do not select:

- *Populations at Risk: Populations at Risk*

[Response Begins]
Elderly (Age >= 65)
[Response Ends]

sp.07. Select the levels of analysis that apply to your measure.

Check ONLY the levels of analysis for which the measure is SPECIFIED and TESTED.

Please refrain from selecting the following answer option(s). We are in the process of phasing out these answer options and request that you instead select one of the other answer options as they apply to your measure.

Please do not select:

- *Clinician: Clinician*
- *Population: Population*

[Response Begins]
Facility
[Response Ends]

sp.08. Indicate the care settings that apply to your measure.

Check ONLY the settings for which the measure is SPECIFIED and TESTED.
[Response Begins]
Post-Acute Care
[Response Ends]

sp.09. 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. If no URL is available, indicate "none available".

[Response Begins]
http://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/NursingHomeQualityInits/NHQIQualityMeasures.html; please see "MDS-3.0-QM-User's-Manual-v14.0.pdf" in the "Users-Manuals-Updated-10-19-2020.zip" zipped folder in the Downloads section at the bottom of the page.
[Response Ends]

sp.11. Attach the data dictionary, code table, or value sets (and risk model codes and coefficients when applicable). Excel formats (.xlsx or .csv) are preferred.

Attach an excel or csv file; if this poses an issue, contact staff. Provide descriptors for any codes. Use one file with multiple worksheets, if needed.
[Response Begins]
No data dictionary/code table – all information provided in the submission form
[Response Ends]

For the question below: state the outcome being measured. Calculation of the risk-adjusted outcome should be described in sp.22.

sp.12. State the numerator.

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.

[Response Begins]
The numerator is the number of long-stay nursing home residents with a selected target assessment indicating a weight loss of 5% or more of the baseline weight in the last 30 days or 10% or more of the baseline weight in the last 6 months who were not on a physician-prescribed weight-loss regimen (K0300 = [2]). The baseline weight is the resident's weight closest to 30 or 180 days before the date of the target assessment.
[Response Ends]

For the question below: describe how the observed outcome is identified/counted. Calculation of the risk-adjusted outcome should be described in sp.22.

sp.13. Provide details needed to calculate the numerator.

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 sp.11.

[Response Begins]
Residents are counted in the numerator if they are long-stay residents, defined as residents who have had 101 or more cumulative days of nursing facility care. Note that the count of cumulative days of nursing facility care continues upon an anticipated reentry within 30 days to the same facility. For example, residents who return to the nursing home following a hospital discharge would not have their length of stay within the episode of care reset to zero if the residents return to the nursing home within 30 days of the prior discharge when return was anticipated. The cumulative days count would resume from the last day of their prior stay. The target population includes all long-stay residents with a target assessment (assessments may be an OBRA quarterly, annual or significant change/correction assessment (A0310A = [02, 03, 04, 05, 06]); or discharge assessment with or without anticipated return (A0310F = [10, 11])), except those with exclusions (specified in sp.16 and sp.17). Note that the PPS assessment schedule changed with the implementation of the Patient Driven Payment Model (PDPM), and PPS 14-, 30-, 60-, and 90-day assessments (A0310B = [02, 03, 04, 05]) are no longer used for target assessments after October 1, 2019. This change may impact the type of target assessment selected for a very small share of long-stay residents who are under SNF care. These residents are still included in the measure denominator, but their target assessment would likely be an OBRA quarterly assessment instead.
The numerator is the number of long-stay residents in the denominator sample with a selected target assessment that indicates a weight loss of 5% or more of the baseline weight in the last month or 10% or more of the baseline weight in the last six months and the resident was not on a physician-prescribed weight loss regimen (K0300=[2]). The baseline weight is the resident's weight closest to 30 or 180 days before the date of the target assessment.
For every calendar quarter (3-month period), the Centers for Medicare & Medicaid Services (CMS) select target assessments conducted during that quarter from each nursing facility to calculate the measure. For any resident with multiple episodes of care during the quarter, only the latest episode will be counted.
A target assessment is defined as the latest assessment that meets the following criteria: (a) it is contained within the resident's selected episode, (b) it has a qualifying reason for assessment, and (c) its target date is no more than 120 days before the end of the episode.

An episode is defined as a period of time spanning one or more stays. An episode begins with an admission and ends with either (a) a discharge, or (b) the end of the target period, whichever comes first. Data are publicly reported on the Nursing Home Compare webpage and are weighted on an average of four target periods.

[Response Ends]

For the question below: state the target population for the outcome. Calculation of the risk-adjusted outcome should be described in sp.22.

sp.14. State the denominator.

Brief, narrative description of the target population being measured.

[Response Begins]
The denominator includes all long-stay residents in the nursing home who have a target assessment (OBRA, PPS or discharge) during the selected quarter and who do not meet the exclusion criteria.
[Response Ends]

For the question below: describe how the target population is identified. Calculation of the risk-adjusted outcome should be described in sp.22.

sp.15. Provide details needed to calculate the denominator.

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 sp.11.

[Response Begins]
Residents are counted in the denominator if they are long-stay residents, defined as residents whose length of stay is 101 days or more. Residents who return to the nursing home following a hospital discharge may not have their length of stay within the episode of care reset to zero. The denominator is the number of long-stay residents with a selected target assessment (assessment types include: OBRA quarterly, annual or significant change/correction assessment (A0310A = [02, 03, 04, 05, 06]); or discharge assessment with or without anticipated return (A0310F = [10, 11])) during the selected quarter, except those with exclusions (specified in sp.16 and sp.17).
[Response Ends]

sp.16. Describe the denominator exclusions.

Brief narrative description of exclusions from the target population.

[Response Begins]
There are four exclusions applied to the denominator: (1) the target assessment is an OBRA admission assessment (A0310A = [01]) or a PPS 5-day assessment (A0310B = [01]), (2) having a prognosis of life expectancy of less than six months (J1400 = [1]) or the six-month prognosis item is missing (J1400 = [-]) on the target assessment, (3) receiving hospice care (O0100K2 = [1]) or the hospice care item is missing (O0100K2 = [-]) on the target assessment, or/and (4) the weight loss item is missing (K0300 = [-]) on the target assessment. Only 1,551 episodes in the 2019 (Q1-Q4) long stay resident sample were excluded from the denominator for this measure due to missing responses on the prognosis of life expectancy being less than 6 months, which accounts for 0.04% of the total episodes. Additionally, only 7,948 (0.241%) episodes in the 2019 (Q1-Q4) long stay residents sample were excluded due to missing responses for the Hospice care item, and only 30,854 (0.935%) episodes were excluded due to missing responses for the weight loss item.

If the facility sample includes fewer than 20 residents after all other resident-level exclusions are applied, then the facility is suppressed from public reporting because of small sample size.
[Response Ends]

sp.17. Provide details needed to calculate the denominator exclusions.

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 sp.11.

[Response Begins]
A long-stay resident is excluded from the denominator if:

1. Target assessment is an OBRA Admission assessment (A0310A= [01]) ***or*** a PPS 5-Day assessment (A0310B= [01])
2. Prognosis of life expectancy is less than 6 months (J1400 = [1]) or the Prognosis item is missing (J1400 = [-]) on the target assessment.
3. Receiving Hospice care (O0100K2 = [1]) or the Hospice care item is missing (O0100K2 = [-]) on the target assessment.
4. Weight loss item is missing (K0300= [-]) on the target assessment.

If the facility sample includes fewer than 20 residents after all other resident-level exclusions are applied, then the facility is suppressed from public reporting because of small sample size.

[Response Ends]

sp.18. Provide all information required to stratify the measure results, if necessary.

Include 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 in the Data Dictionary field.

[Response Begins]
This measure is not stratified.
[Response Ends]

sp.19. Select the risk adjustment type.

Select type. Provide specifications for risk stratification and/or risk models in the Scientific Acceptability section.
[Response Begins]
No risk adjustment or risk stratification
[Response Ends]

sp.20. Select the most relevant type of score.

Attachment: If available, please provide a sample report.
[Response Begins]
Rate/proportion
[Response Ends]

sp.21. Select the appropriate interpretation of the measure score.

Classifies interpretation of score according to whether better quality or resource use is associated with a higher score, a lower score, a score falling within a defined interval, or a passing score
[Response Begins]
Better quality = Lower score
[Response Ends]

sp.22. Diagram or describe the calculation of the measure score as an ordered sequence of steps.

Identify the target population; exclusions; cases meeting the target process, condition, event, or outcome; time period of data, aggregating data; risk adjustment; etc.

[Response Begins]
The Percent of Residents Who Lose Too Much Weight (NQF 0689) is primarily publicly reported as a four-quarter measure, which is based on a rolling four-quarter weighted average that is updated quarterly on Care Compare (https://www.medicare.gov/care-compare/). The four-quarter measure score is computed as follows:

$$QM_{4Quater} = [(QM_{Q1} * D_{Q1}) + (QM_{Q2} * D_{Q2}) + (QM_{Q3} * D_{Q3}) + QM_{Q4} * D_{Q4})]/(D_{Q1} + D_{Q2} + D_{Q3} + D_{Q4})$$

Where QM_{Q1}, QM_{Q2}, QM_{Q3}, and QM_{Q4} correspond to the QM values for the four quarters, and D_{Q1}, D_{Q2}, D_{Q3} and D_{Q4} are the denominators (number of eligible residents) for the four quarters.

Outlined below are the steps for calculating the quarterly score for this measure.

Step 1: Identify the total number of long-stay residents who have a target assessment (OBRA, PPS, or discharge) during quarter and don't meet the exclusion criteria.

Step 2: Starting with the set of residents identified in Step 1, determine the number of long-stay residents who have experienced weight loss of 5% or more in the last month or 10% or more in the last six months and the weight loss was not planned or prescribed by a physician (K0300=[02]).

Step 3: Divide the result of Step 2 by the result of Step 1.

Step 4: Multiply the result of step 3 by 100 to obtain a percent value.

A description of the time period for the data included in this measure is provided in sp.13 above.

[Response Ends]

sp.25. If measure is based on a sample, provide instructions for obtaining the sample and guidance on minimum sample size.

[Response Begins]
This is not applicable because the data are not estimated based on samples. Rather, the data include all nursing home residents nationally who do not meet the exclusion criteria.
[Response Ends]

sp.28. Select only the data sources for which the measure is specified.

[Response Begins]
Assessment Data
[Response Ends]

sp.29. Identify the specific data source or data collection instrument.

For example, provide the name of the database, clinical registry, collection instrument, etc., and describe how data are collected.

[Response Begins]
The data source is the Minimum Data Set (MDS) 3.0, and the collection instrument is the Resident Assessment Instrument (RAI). For MDS 3.0 item sets used to calculate the quality measure, please see "MDS3.0_Final_Item_Sets_v1.17.2 for October 1 2020 zip (ZIP)" under the "Downloads" section of the following webpage: https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/NursingHomeQualityInits/NHQIMDS30TechnicalInformation
[Response Ends]

sp.30. Provide the data collection instrument.

[Response Begins]
Available at measure-specific web page URL identified in sp.09
[Response Ends]

2ma.01. Indicate whether additional empirical reliability testing at the accountable entity level has been conducted. If yes, please provide results in the following section, Scientific Acceptability: Reliability - Testing. Include information on all testing conducted (prior testing as well as any new testing); use red font to indicate updated testing.

[Response Begins]
Yes
[Response Ends]

2ma.02. Indicate whether additional empirical validity testing at the accountable entity level has been conducted. If yes, please provide results in the following section, Scientific Acceptability: Validity - Testing. Include information on all testing conducted (prior testing as well as any new testing); use red font to indicate updated testing.

[Response Begins]
Yes
[Response Ends]

2ma.03. For outcome, patient-reported outcome, resource use, cost, and some process measures, risk adjustment/stratification may be conducted. Did you perform a risk adjustment or stratification analysis?

[Response Begins]
Yes
[Response Ends]

2ma.04. For maintenance measures in which risk adjustment/stratification has been performed, indicate whether additional risk adjustment testing has been conducted since the most recent maintenance evaluation. This may include updates to the risk adjustment analysis with additional clinical, demographic, and social risk factors.

Please update the Scientific Acceptability: Validity - Other Threats to Validity section.

Note: This section must be updated even if social risk factors are not included in the risk adjustment strategy.

[Response Begins]
Yes - Additional risk adjustment analysis is included
[Response Ends]

Measure testing must demonstrate adequate reliability and validity in order to be recommended for endorsement. Testing may be conducted for data elements and/or the computed measure score. Testing information and results should be entered in the appropriate fields in the Scientific Acceptability sections of the Measure Submission Form.

- Measures must be tested for all the data sources and levels of analyses that are specified. If there is more than one set of data specifications or more than one level of analysis, contact NQF staff about how to present all the testing information in one form.
- All required sections must be completed.
- For composites with outcome and resource use measures, Questions 2b.23-2b.37 (Risk Adjustment) also must be completed.
- If specified for multiple data sources/sets of specifications (e.g., claims and EHRs), Questions 2b.11-2b.13 also must be completed.
- An appendix for supplemental materials may be submitted (see Question 1 in the Additional section), but there is no guarantee it will be reviewed.
- Contact NQF staff with any questions. Check for resources at the Submitting Standards webpage.
- For information on the most updated guidance on how to address social risk factors variables and testing in this form refer to the release notes for the 2021 Measure Evaluation Criteria and Guidance.

Note: The information provided in this form is intended to aid the Standing Committee and other stakeholders in understanding to what degree the testing results for this measure meet NQF's evaluation criteria for testing.
2a. Reliability testing demonstrates 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. For instrument-based measures (including PRO-PMs) and composite performance measures, reliability should be demonstrated for the computed performance score.
2b1. Validity testing demonstrates that the measure data elements are correct and/or the measure score correctly reflects the quality of care provided, adequately identifying differences in quality. For instrument based measures (including PRO-PMs) and composite performance measures, validity should be demonstrated for the computed performance score.
2b2. Exclusions are supported by the clinical evidence and are of sufficient frequency to warrant inclusion in the specifications of the measure;
AND
If patient preference (e.g., informed decision-making) is a basis for exclusion, there must be evidence that the exclusion impacts performance on the measure; in such cases, the measure must be specified so that the information about patient preference and the effect on the measure is transparent (e.g., numerator category computed separately, denominator exclusion category computed separately).
2b3. For outcome measures and other measures when indicated (e.g., resource use):

- an evidence-based risk-adjustment strategy (e.g., risk models, risk stratification) is specified; is based on patient factors (including clinical and social risk factors) that influence the measured outcome and are present at start of care; 14,15 and has demonstrated adequate discrimination and calibration

OR

- rationale/data support no risk adjustment/ stratification.

2b4. Data analysis of computed measure scores demonstrates that methods for scoring and analysis of the specified measure allow for identification of statistically significant and practically/clinically meaningful 16 differences in performance;
OR
there is evidence of overall less-than-optimal performance.
2b5. If multiple data sources/methods are specified, there is demonstration they produce comparable results.
2b6. Analyses 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 non-responders) and how the specified handling of missing data minimizes bias.
2c. For composite performance measures, empirical analyses support the composite construction approach and demonstrate that:
2c1. the component measures fit the quality construct and add value to the overall composite while achieving the related objective of parsimony to the extent possible; and
2c2. the aggregation and weighting rules are consistent with the quality construct and rationale while achieving the related objective of simplicity to the extent possible.

(if not conducted or results not adequate, justification must be submitted and accepted)

Definitions
Reliability testing applies to both the data elements and computed measure score. Examples of reliability testing for data elements include, but are not limited to: inter-rater/abstractor or intra-rater/abstractor studies; internal consistency for multi-item scales; test-retest for survey items. Reliability testing of the measure score addresses precision of measurement (e.g., signal-to-noise).
Validity testing applies to both the data elements and computed measure score. Validity testing of data elements typically analyzes agreement with another authoritative source of the same information. Examples of validity testing of the measure score include, but are not limited to: testing hypotheses that the measures scores indicate quality of care, e.g., measure scores are different for groups known to have differences in quality assessed by another valid quality measure or method; correlation of measure scores with another valid indicator of quality for the specific topic; or relationship to conceptually related measures (e.g., scores on process measures to scores on outcome measures). Face validity of the measure score as a quality indicator may be adequate if accomplished through a systematic and transparent process, by identified experts, and explicitly addresses whether performance scores resulting from the measure as specified can be used to distinguish good from poor quality. The degree of consensus and any areas of disagreement must be provided/discussed.
Examples of evidence that an exclusion distorts measure results include, but are not limited to: frequency of occurrence, variability of exclusions across providers, and sensitivity analyses with and without the exclusion.
Patient preference is not a clinical exception to eligibility and can be influenced by provider interventions.
Risk factors that influence outcomes should not be specified as exclusions.
With large enough sample sizes, small differences that are statistically significant may or may not be practically or clinically meaningful. The substantive question may be, for example, whether a statistically significant difference of one percentage point in the percentage of patients who received smoking cessation counseling (e.g., 74 percent v. 75 percent) is clinically meaningful; or whether a statistically significant difference of $25 in cost for an episode of care (e.g., $5,000 v.$5,025) is practically meaningful. Measures with overall less-than-optimal performance may not demonstrate much variability across providers.

For more details on updated testing of the weight loss measure, please see the red text between the [Response begins] and [Response Ends] notifications in Sections 2a and 2b.

2a. Reliability

2a.01. Select only the data sources for which the measure is tested.

[Response Begins]
Assessment Data
[Response Ends]

2a.02. 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).

[Response Begins]
[2021 submission] The dataset used for testing was the Nursing Home Minimum Data Set (MDS) 3.0, which is one of three components of the Resident Assessment Instrument (RAI). The RAI is a tool used by nursing home staff to collect information on residents' strengths and needs. The MDS contains screening, clinical, and functional status elements, such as definitions and coding categories. These elements form the foundation of the comprehensive RAI for all eligible Medicare and Medicaid beneficiaries who are residents of nursing homes. The MDS items standardize how information about resident status and condition is recorded and shared within the facility, between facilities, and between facilities and outside agencies. Nursing homes are required to complete assessments on a regular basis, and the assessment

requirements for the MDS are applicable to all residents in Medicare and/or Medicaid certified long-term care facilities, regardless of payment source or payer source.
[Response Ends]

2a.03. Provide the dates of the data used in testing.

Use the following format: "MM-DD-YYYY - MM-DD-YYYY"

[Response Begins]
Critical Data Element Testing (08-01-2006 - 02-28-2007)

[2015 submission] The RAND Development and Validation study from August 2006 to February 2007 on the development and validation of a revised nursing home assessment tool was used for the testing of critical data elements. [2021 submission] It is important to note that the MDS 3.0 forms used in the RAND study are similar to the MDS 3.0 v1.17.0 forms used in the testing of this measure. The MDS 3.0 item set has remained stable since RAND created the recommended MDS 3.0 form in 2008, with the exception of select changes unrelated to this measure (changes in item specifications and the addition of some new items). In particular, the weight loss item has the same item wording in the MDS 3.0 v1.17.0 and the 2008 recommended form.

[2015 submission] Saliba, D., & Buchanan, J. (2008, April). Development and validation of a revised nursing home assessment tool: MDS 3.0. Contract No. 500-00-0027/Task Order #2. Santa Monica, CA: Rand Corporation. Retrieved from https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/NursingHomeQualityInits/downloads/MDS30FinalReport.pdf.

[2021 submission] The authors of the RAND study also conducted an evaluation of the MDS 3.0 form in 2012 to determine whether their revisions improved reliability, validity, resident input, and clinical utility, all while decreasing collection burden. The results of this 2012 follow-up study were also used for the testing of critical data elements.

Saliba, D., & Buchanan, J. (2012). "Making the Investment Count: Revision of the Minimum Data Set for Nursing Homes, MDS 3.0." *Journal of American Medical Directors Association* 13(7): 602-10. https://doi.org/10.1016/j.jamda.2012.06.002.

Performance Measure Score Testing (01-01-2019 - 12-31-2019)

Since NQF #0689, Percent of Residents Who Lose Too Much Weight, is primarily publicly reported by a four-quarter weighted average score, four quarters of MDS 3.0 data (2019Q1-2019Q4) were used to construct this measure and calculate the QM scores.

[Response Ends]

2a.04. Select the levels of analysis for which the measure is tested.

Testing must be provided for all the levels specified and intended for measure implementation, e.g., individual clinician, hospital, health plan.

Please refrain from selecting the following answer option(s). We are in the process of phasing out these answer options and request that you instead select one of the other answer options as they apply to your measure.

Please do not select:

- *Clinician: Clinician*
- *Population: Population*

[Response Begins]
Facility

[Response Ends]

2a.05. List the measured entities 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.

[Response Begins]

[2015 submission] ***Critical Data Element Testing***

The RAND Development and Validation of MDS 3.0 study sample included a representative sample of for-profit and not-for-profit facilities, and hospital-based and freestanding facilities, which were recruited for the study. The sample included 71 community nursing facilities in 8 states and 19 Veterans Affairs (VA) nursing homes. Approximately 63% of the 71 community facilities were for profit, 35% were not-for-profit, and 1% were government owned. Most facilities (90%) were free-standing nursing homes and 10% were hospital-based. Over half of the sample facilities (57%) had 100-199 beds, 22% had over 200 beds, and 4% had fewer than 50. Most nursing homes (78%) were located in urban areas (Saliba & Buchanan, 2008).

Saliba, D., & Buchanan, J. (2008, April). Development and validation of a revised nursing home assessment tool: MDS 3.0. Contract No. 500-00-0027/Task Order #2. Santa Monica, CA: Rand Corporation. Retrieved from https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/NursingHomeQualityInits/downloads/MDS30FinalReport.pdf.

[2021 submission] **Performance Measure Score Testing**

The analysis of MDS 3.0 data included all nationwide nursing home facilities with sufficient denominator size (n ≥ 20) to publicly report this measure between 2019Q1 and 2019Q4 (k = 14,829), unless otherwise noted. Active diagnoses and item I8000 of the MDS 3.0 are not used in the specification of this measure. Some MDS active diagnosis checkbox items between I0100 - I6500 are used in Table 1 in section 2a.06) below for a display of population characteristics.

[Response Ends]

2a.06. Identify the number and descriptive characteristics of patients included in the analysis (e.g., age, sex, race, diagnosis), separated by level of analysis and data source; if a sample was used, describe how patients were selected for inclusion in the sample.

If there is a minimum case count used for testing, that minimum must be reflected in the specifications.

[Response Begins]

[2015 submission] ***Critical Data Element Testing***

The RAND Development and Validation of MDS 3.0 study sample included 3,822 residents from 71 community nursing homes and 764 residents from 19 VHA nursing homes (Saliba & Buchanan, 2008).

Saliba, D., & Buchanan, J. (2008, April). Development and validation of a revised nursing home assessment tool: MDS 3.0. Contract No. 500-00-0027/Task Order #2. Santa Monica, CA: Rand Corporation. Retrieved from https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/NursingHomeQualityInits/downloads/MDS30FinalReport.pdf.

[2021 submission] **Performance Measure Score Testing**

This measure is for residents who lose too much weight in nursing facilities, which is not the result of a physician-prescribed weight-loss regimen. The analysis of MDS 3.0 data included all long-stay residents who met the denominator inclusion criteria for this measure in facilities with sufficient sample size (n ≥ 20, k = 14,274) and reported this measure between 2019Q1 and 2019Q4. 932,094 residents met the denominator inclusion criteria in these facilities.

Table 1 describes the characteristics of the residents who were counted in the denominator after applying facility sample size restrictions to 2019Q4 data (n = 932,094). The majority of residents who met the denominator criteria were female (63.2%) and white (72.1%), while a smaller proportion of residents were male (36.8%) and Black or African American (16.7%). A majority of residents were dual-eligible for Medicare and Medicaid (73.7%). More than 36% of residents were over the age of 85, and approximately 26% were between the ages of 75-84. The most frequently reported diagnoses were Hypertension (77.9%), Depression (52.8%), and Non-Alzheimer's Dementia (48.4%). Other common diagnoses reported for more than a quarter of residents were Arthritis (32.1%), Anemia

(30.8%), Diabetes Mellitus (36.0%), and Anxiety Disorders (31.5%). Table 1 also outlines the characteristics of the residents who were counted in the numerator. Compared to the denominator, the numerator had a higher share of females, Medicare only residents, residents aged 75 or older, residents with hip fractures and other fracture, and residents with hip fractures, other fractures, Non-Alzheimer's Dementia, malnutrition, and cancer. These trends observed in **Table 1** are consistent across all other quarters that were tested (2019Q3, 2019Q2, and 2019Q1).

Table 1. Characteristics of Long-Stay Residents Included in Analyses, NQF #0689 (2019Q4)

*	NQF #0689 Denominator	NQF #0689 Denominator	NQF #0689 Denominator	NQF #0689 Numerator	NQF #0689 Numerator	NQF #0689 Numerator	Percentage Ratio (% Numerator/ % Denominator)
Resident characteristics	**Frequency (n)**	**Total Observations (N)**	**Percentage (%)**	**Frequency (n)**	**Total Observations (N)**	**Percentage (%)**	*
Sex	**No data**	No data	No data	No data	No data	No data	No data
Female	588,808	932,094	63.2	32,535	49,919	65.2	1.03
Male	343,286	932,094	36.8	17,384	49,919	34.8	0.95
Race/Ethnicity	**No data**	No data	No data	No data	No data	No data	No data
White Only	672,425	932,094	72.1	36,336	49,919	72.8	1.01
Black or African American Only	155,393	932,094	16.7	8,358	49,919	16.7	1.00
Hispanic or Latino Only	55,794	932,094	6.0	2,681	49,919	5.4	0.90
Asian Only	21,484	932,094	2.3	1,101	49,919	2.2	0.96
American Indian/Alaska Native Only	4,582	932,094	0.5	218	49,919	0.4	0.89
Native Hawaiian or Other Pacific Islander Only	3,514	932,094	0.4	167	49,919	0.3	0.89

*	NQF #0689 Denominator	NQF #0689 Denominator	NQF #0689 Denominator	NQF #0689 Numerator	NQF #0689 Numerator	NQF #0689 Numerator	Percentage Ratio (% Numerator/ % Denominator)
Multi-race	3,007	932,094	0.3	148	49,919	0.3	0.92
Medicare-Medicaid Dual Eligibility	**No data**	No data	No data	No data	No data	No data	No data
Medicaid & Medicare	686,984	932,094	73.7	36,317	49,919	72.8	0.99
Medicaid Only	109,654	932,094	11.8	5,109	49,919	10.2	0.87
Medicare Only	121,586	932,094	13.0	7,643	49,919	15.3	1.17
Neither Medicaid nor Medicare	13,870	932,094	1.5	850	49,919	1.7	1.14
Age	**No data**	No data	No data	No data	No data	No data	No data
<65	161,514	932,094	17.3	7,048	49,919	14.1	0.81
65-74	184,478	932,094	19.8	9,169	49,919	18.4	0.93
75-84	245,660	932,094	26.4	13,690	49,919	27.4	1.04
85+	340,442	932,094	36.5	20,012	49,919	40.1	1.10
Diagnoses	**No data**	No data	No data	No data	No data	No data	No data
Arthritis	73,116	228,088	32.1	4,551	13,962	32.6	1.02
Osteoporosis	28,827	228,086	12.6	1,826	13,960	13.1	1.03
Hip Fracture	14,451	848,860	1.7	1,312	42,787	3.1	1.80
Other Fracture	29,471	848,853	3.5	1,987	42,788	4.6	1.34
Depression	448,225	848,817	52.8	23,660	42,787	55.3	1.05

*	NQF #0689 Denominator	NQF #0689 Denominator	NQF #0689 Denominator	NQF #0689 Numerator	NQF #0689 Numerator	NQF #0689 Numerator	Percentage Ratio (% Numerator/ % Denominator)
Stroke	143,467	848,843	16.9	7,450	42,787	17.4	1.03
Alzheimer's Disease	125,939	848,840	14.8	7,214	42,788	16.9	1.14
Non-Alzheimer's Dementia	410,539	848,805	48.4	22,725	42,788	53.1	1.10
Malnutrition or at risk for malnutrition	59,062	932,056	6.3	5,445	49,916	10.9	1.72
Cancer	51,671	839,302	6.2	3,088	41,719	7.4	1.20
Anemia	261,500	848,819	30.8	14,436	42,785	33.7	1.10
Heart Failure	178,799	848,845	21.1	10,045	42,788	23.5	1.11
Hypertension	661,555	848,815	77.9	33,406	42,784	78.1	1.00
Diabetes Mellitus	335,504	932,078	36.0	18,248	49,918	36.6	1.02
Anxiety Disorder	293,893	932,024	31.5	16,609	49,915	33.3	1.06
Asthma, Chronic Obstructive Pulmonary Disease, or Chronic Lung Disease	201,179	848,842	23.7	10,556	42,788	24.7	1.04

***This cell is intentionally left empty.**

[Response Ends]

2a.07. 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.

[Response Begins]
[2015 submission] **Data for Critical Data Elements**
RAND reliability analysis of data elements used the same sample as described in ***Sections 2a.05*** *and* ***2a.06*** *(Saliba & Buchanan, 2008).*

Saliba, D., & Buchanan, J. (2008, April). Development and validation of a revised nursing home assessment tool: MDS 3.0. Contract No. 500-00-0027/Task Order #2. Santa Monica, CA: Rand Corporation. Retrieved from https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/NursingHomeQualityInits/downloads/MDS30FinalReport.pdf.

[2021 submission] **Data for Measure Performance Score Testing**

All analyses used the same data as described above in **Sections 2a.02, 2a.03, and 2a.05**.

[Response Ends]

2a.08. List 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.

[Response Begins]
[2021 submission] Resident-level social risk factor variables related to weight loss that were available in the MDS 3.0 dataset were selected, including age, race, Medicaid status, and gender. The descriptive statistics for all of these characteristics are listed in Table 1 under item 2a.06.
[Response Ends]

Note: If accuracy/correctness (validity) of data elements was empirically tested, separate reliability testing of data elements is not required – in 2a.07 check patient or encounter-level data; in 2a.08 enter "see validity testing section of data elements"; and enter "N/A" for 2a.09 and 2a.10.

2a.09. Select the level of reliability testing conducted.

Choose one or both levels.
[Response Begins]
Patient or Encounter-Level (e.g., inter-abstractor reliability; data element reliability must address ALL critical data elements)
Accountable Entity Level (e.g., signal-to-noise analysis)
[Response Ends]

2a.10. For each level of reliability testing 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.

[Response Begins]
[2015 submission] ***Critical Data Element Reliability***

1. *The national test of MDS 3.0 items examined the agreement between assessors (reliability). Inter-rater reliability measures the extent to which two data collectors achieve the same results when assessing the same resident within the same time frame. Two types of reliability were tested: gold-standard nurse to gold-standard nurse, and gold-standard nurse to facility-nurse. Quality Improvement Organizations were employed to identify gold-standard (research) nurses and recruit community nursing facilities to participate in the national evaluation (Saliba & Buchanan, 2008). The gold-standard nurses were trained in the MDS 3.0 instrument, and they, in turn, trained a facility nurse from each participating nursing facility in their home states. The gold-standard to gold-standard comparisons provided information on instrument performance with highly trained nurses using research*

protocols. The gold-standard to facility-nurse comparisons measured item performance in a more operational environment in which one assessor had ongoing facility responsibilities. Residents participating in the test were selected to capture a representative sample of short- and long-stay residents. Kappa statistics were calculated to assess item reliability. Kappa is a statistical measure of inter-rater agreement for qualitative data, ranging from 0.0 to 1.0. A rating of 0.70 is considered substantial agreement.

Landis, JR, Koch, GG. The measurement of observer agreement for categorical data. Biometrics 33(1), p 159-174, 1977.
Saliba, D., & Buchanan, J. (2008, April). Development and validation of a revised nursing home assessment tool: MDS 3.0. Contract No. 500-00-0027/Task Order #2. Santa Monica, CA: Rand Corporation. Retrieved from https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/NursingHomeQualityInits/downloads/MDS30FinalReport.pdf.

[2021 submission] Performance Measure Score Reliability

1. Split-half reliability analysis: Split-half reliability assesses the internal consistency of a quality measure by randomly dividing the residents within each nursing facility into two halves and calculating the correlation between each facility's quality measure scores on the basis of the two divided halves. When a nursing facility's residents, randomly divided, have a score similar to one another, the quality measure score is more likely to reflect systematic differences in nursing home-level quality rather than random variation. In this analysis, a split-half reliability analysis was conducted on all facilities with 40 or more residents counted in the measure denominator across the eight quarters (ensuring at least 20 residents could be used in each randomly selected half of a facility's residents). Data from 2018Q1 -2019Q4 were used to calculate the Spearman Rank Correlation and Pearson Correlation to measure the internal reliability.
2. Signal-to-noise analysis: The signal-to-noise ratio gives the proportion of variability in measure performance that can be explained by between-provider differences in provider performance rather than variability within a provider (e.g. through measurement or sampling error). Since experiencing weight-loss is a binary outcome, the reliability was estimated using a beta-binomial model. The beta-binomial model assumes that the provider QM score for the weight loss measure is a binomial random variable, conditional on the provider's true value that comes from a beta distribution. Data from 2019Q1 through 2019Q4 were used to conduct this analysis by fitting the beta binomial model to the data. The estimated alpha and beta parameters from the model were used to calculate the provider-to-provider variance:

$$\sigma^2_{\text{provider-to-provider}} = \frac{\alpha\beta}{(\alpha+\beta+1)(\alpha+\beta)^2}$$

The provider-specific error was calculated using the following formula, where "p" is each facility's QM score and "n" is the number of residents in each facility:

$$\sigma^2_{\text{provider-to-provider}} = \frac{\alpha\beta}{(\alpha+\beta+1)(\alpha+\beta)^2}$$

The reliability score for each facility was then calculated using the following formula:

$$\text{reliability} = \frac{\sigma^2_{\text{provider-to-provider}}}{\sigma^2_{\text{provider-to-provider}} + \sigma^2_{\text{provider-specific-error}}}$$

A reliability score closer to 1 implies that most of the variability is attributable to between-provider differences in performance, and a score closer to 0 implies that most of the variability in the measure is attributable to variation within providers.

[Response Ends]

2a.11. For each level of reliability testing checked above, what were the statistical results from reliability testing?

For example, provide the percent agreement and kappa for the critical data elements, or distribution of reliability statistics from a signal-to-noise analysis. For score-level reliability testing, when using a signal-to-noise analysis, more than just one overall statistic should be reported (i.e., to demonstrate variation in reliability across providers). If a particular method yields only one statistic, this should be explained. In addition, reporting of results stratified by sample size is preferred (pg. 18, NQF Measure Evaluation Criteria).

[Response Begins]

[2015 submission] ***Critical Data Element Reliability***

1. *In their testing of the MDS 3.0, RAND observed that the kappa for gold-standard to gold-standard assessments of the weight loss item was 0.944; and the kappa for gold-standard nurse assessment to facility nurse assessment of weight loss item was 0.918. The kappa for gold-standard to gold-standard assessments of the 6-month prognosis item was 0.8724; and the kappa for gold-standard nurse assessment to facility nurse assessment of the 6-month prognosis item was 0.964. Kappa is a statistical measure of inter-rater agreement for qualitative data, ranging from 0.0 to 1.0. A rating of 0.964 is considered "substantial agreement." These results are indicative of data element reliability.*

   Saliba, D., & Buchanan, J. (2008, April). Development and validation of a revised nursing home assessment tool: MDS 3.0. Contract No. 500-00-0027/Task Order #2. Santa Monica, CA: Rand Corporation. Retrieved from https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/NursingHomeQualityInits/downloads/MDS30FinalReport.pdf.

   [2021 submission] **Performance Measure Score Reliability**

1. Split-half reliability analysis: The split-half correlation for this measure was positive, and the relationship was moderate (r = 0.64, ρ = 0.65, ICC = 0.64, p < .01), suggesting there is considerable evidence of internal reliability. These moderate correlations were expected due to a modest amount of variation in performance among providers. Since correlations are calculated using the covariance of the data and the individual variances to naturalize the covariance to report a value range between -1 and 1, the modest amount of variance in performance was expected to yield moderate correlation coefficients. **Table 6** in **Section 2b4.2** demonstrates that the variation in scores is still sufficient to distinguish high-performers and low-performers.
2. Signal-to-noise analysis: The average signal-to-noise reliability score of this quality measure using facility scores based on 2019Q1-2019Q4 data was observed to be 0.76. This suggests that the measure is very reliable in separating facility characteristics from variability within facility. This moderate variability is expected for this outcome measure because of modest variability in measure scores.

[Response Ends]

2a.12. Interpret the results, in terms of how they demonstrate reliability.

(In other words, what do the results mean and what are the norms for the test conducted?)

[Response Begins]

[2015 submission] **Critical Data Element Reliability**

The RAND Development and Validation of MDS 3.0 national pilot test study demonstrated excellent reliability for MDS 3.0 items used to calculate this measure. Although the RAND testing was conducted 13 years ago, the MDS 3.0 forms used in the RAND study are similar to the latest MDS 3.0 forms used in the testing of this measure. The MDS 3.0 item set has remained stable since RAND created the recommended MDS 3.0 form in 2008, with the exception of select changes in item specifications and the addition of some new items. In particular, the Weight Loss item has the same item wording in the latest MDS 3.0 form and the 2008 recommended form.

[2021 submission] **Performance Measure Score Reliability**

These analyses demonstrate that the weight loss measure shows considerable evidence of internal reliability. The average signal-to-noise ratio across all providers was 0.76, meaning 76% of the variance in scores for this measure were explained by inter-facility variation. This suggests that the measure is very reliable in separating provider characteristics from variability within provider.

[Response Ends]

Validity Testing

2b.01. Select the level of validity testing that was conducted.

[Response Begins]
Patient or Encounter-Level (data element validity must address ALL critical data elements)
Accountable Entity Level (e.g. hospitals, clinicians)
Empirical validity testing
[Response Ends]

2b.02. 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.

[Response Begins]
Critical Data Element Validity

[2015 submission] The RAND validation of MDS 3.0 study tested the criterion validity of the items by comparing how different nurses assessed the same residents using MDS 3.0. They compared gold-standard research nurses to gold-standard nurses, and they compare gold-standard nurses to staff nurses trained by the gold-standard nurses. Kappa statistic was calculated.

[2021 submission] **Performance Measure Score Validity**

1. Convergent validity: Groups of quality measures that reflect similar care processes or outcomes were examined with the hypothesis that a facility's percentile ranking (compared to all facilities reporting the measure) may be somewhat consistent among related quality measures. Related MDS Quality Measures (Percent of Residents Whose Ability to Move Independently Worsened, Percent of Residents Whose Need for Help with Activities of Daily Living Has Increased, and Percent of Residents Who Have Depressive Symptoms) that are associated with the risk of weight-loss that is not related to a physician-prescribed weight-loss regimen and Facility Five-Star Ratings were examined for this purpose. Public reporting data was used to calculate these correlations between NQF #0689 (Percent of Residents Who Lose Too Much Weight (long stay)) and related quality measures.
2. Variation by state: Analyses investigated whether or not variation in scores on this measure was substantially attributable to state-by-state differences. If a measure is subject to variation caused by other factors beyond facility control, such as state-level payment policies or demographics, this variation can be a threat to the validity of the measure. At the same time, it is expected that state variation may explain a small portion of measure variation due to differences in quality across states.
3. Seasonality: Another potential threat to the validity of a quality measure is seasonal variation. If a quality measure score varies substantially from quarter to quarter in a consistent pattern over time corresponding to changes in seasons, it is possible that the validity of the measure is being compromised due to influences not within a nursing home's control. To address whether seasonal variation might play a role, the trend in the national mean and median for this quality measure score between 2018Q1 and 2019Q4 was examined.
4. Stability analysis: The extent to which relative facility rank changed on this quality measure from one reporting period to another was also assessed by evaluating the percentage of facilities that changed in their percentile ranking (i.e., relative quality measure score) within 1 decile, 2 deciles, and 3 or more deciles. Dramatic changes in the quality measure score or facility rank based on the score over time may indicate measure instability, rather than true changes in quality. An important caveat is that some degree of variation in performance across time is to be expected: very poor performance in one quarter may lead to immediate changes that improve performance in subsequent quarters, and some movement in performance becomes more likely with rare event outcomes.
5. Confidence interval analysis: Proportions of facilities with scores for this measure that are significantly different from the national facility-level mean were examined and stratified by facility denominator size. For this analysis, statistical significance was determined by using 95% confidence intervals. A facility's quality measure score was significantly different from the national mean if the national mean was not included in the facility's 95% confidence interval. High-performing facilities should have scores that are significantly better than average, and

scores of low-performing facilities should be significantly below average. The analysis was stratified by facility denominator size to examine whether this feature of the measure varies by size.

[Response Ends]

2b.03. Provide the statistical results from validity testing.

Examples may include correlations or t-test results.

[Response Begins]
Critical Data Elements

[2015 submission] The kappa for gold-standard nurse assessment to facility nurse assessment of weight loss item was 0.918. The kappa for gold-standard nurse assessment to facility nurse assessment of the 6-month prognosis item was 0.964.

References:

1. *Saliba, D., & Buchanan, J. (2008, April). Development and validation of a revised nursing home assessment tool: MDS 3.0. Contract No. 500-00-0027/Task Order #2. Santa Monica, CA: Rand Corporation. Retrieved from https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/NursingHomeQualityInits/downloads/MDS30FinalReport.pdf.*

[2021 submission] **Performance Measure Score Validity**

- Convergent Validity: Among facilities that could report all measures, the analysis found statistically significant positive correlation between Percent of Residents Who Lose Too Much Weight (NQF #0689) and Percent of Residents Whose Ability to Move Independently Worsened, Percent of Residents Whose Need for Help with Activities of Daily Living Has Increased, and Percent of Residents Who Have Depressive Symptoms. Statistically significant negative correlations between Percent of Residents Who Lose Too Much Weight and Overall Facility Five-Star Ratings, Quality Ratings, Staffing Ratings, and Registered Nurse Staffing Ratings were also observed. The coefficient estimates and associated p-values are summarized in **Table 2** below.

Table 2. Correlations between NQF #0689 and other related MDS Quality Measures, Facility Five-Star Ratings and Claims-based Quality Measures (2019Q1-2019Q4)

Quality Measure	Spearman Correlation	P-Value
MDS Quality Measures	**No data**	**No data**
Percent of Residents Whose Ability to Move Independently Worsened (Long Stay)	0.113	<.0001
Percent of Residents Whose Need for Help with Activities of Daily Living Has Increased (Long Stay)	0.108	<.0001
Percent of Residents Who Have Depressive Symptoms (Long Stay)	0.063	<.0001
Facility Five-Star Ratings	**No data**	**No data**
Overall facility ratings	-0.108	<.0001
Quality ratings	-0.143	<.0001
Staffing ratings	-0.029	<.0001
Registered nurse staffing ratings	-0.011	<.0001

- Variation by State: The proportion of variation in this measure explained by the state that facilities are located in was small though significant ($p < .001$). An analysis of variance showed that just 4.8% of the overall variance in this measure can be attributed to the state in which the facility is located. The average inter-quartile range of state-level scores was 3.8 percentage points. Alaska had the highest mean for NQF #0689 compared to the other

states but had a relatively low number of nursing facilities located in the state. The state-level average scores and percentile distributions are summarized in **Table 3** below.

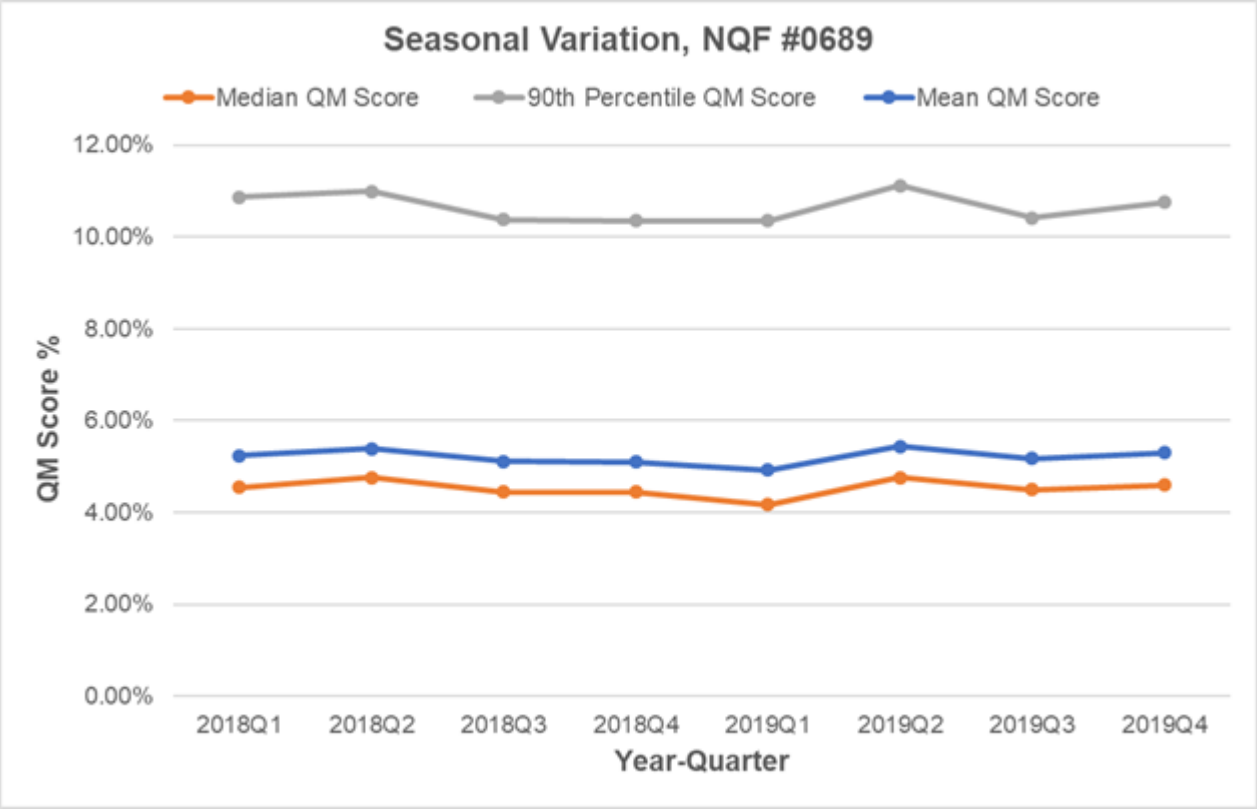
Table 3. State-level NQF #0689 QM score summary (2019Q1-2019Q4)

State	Number of facilities	Mean score	Std dev.	10th percentile	25th percentile	50th percentile	75th percentile	90th percentile	Interquartile range
AK	8	7.3%	3.6%	3.6%	4.5%	6.5%	9.1%	14.5%	4.6%
WV	113	6.8%	3.2%	2.9%	4.5%	6.7%	8.6%	10.8%	4.1%
SC	159	6.7%	3.9%	2.9%	4.1%	6.2%	8.5%	11.0%	4.3%
NC	398	6.7%	3.2%	2.8%	4.3%	6.3%	8.7%	11.0%	4.5%
VT	33	6.6%	3.3%	2.5%	4.5%	5.8%	9.3%	10.2%	4.8%
KY	263	6.4%	3.5%	2.9%	4.2%	5.9%	8.2%	10.3%	4.0%
SD	104	6.3%	3.9%	1.6%	3.7%	5.6%	7.9%	11.8%	4.2%
WY	36	6.2%	3.5%	2.5%	3.6%	5.6%	8.3%	11.4%	4.7%
TN	294	6.2%	3.5%	2.3%	3.9%	5.6%	8.1%	10.5%	4.2%
MT	69	6.0%	3.5%	2.2%	4.0%	5.4%	7.8%	9.2%	3.8%
IL	655	5.9%	3.9%	1.4%	3.2%	5.5%	8.1%	10.8%	4.8%
GA	350	5.8%	3.1%	1.9%	3.7%	5.5%	7.7%	9.8%	4.0%
MS	198	5.8%	3.1%	1.6%	3.8%	5.6%	7.5%	10.2%	3.7%
AL	223	5.7%	2.8%	2.4%	4.0%	5.5%	7.1%	9.4%	3.1%
VA	265	5.7%	2.8%	2.3%	3.6%	5.5%	7.4%	9.1%	3.9%
PA	647	5.6%	2.7%	2.4%	3.7%	5.2%	7.2%	9.1%	3.5%
NY	598	5.6%	2.7%	2.5%	3.8%	5.4%	7.1%	9.0%	3.3%
IN	513	5.6%	3.2%	1.9%	3.3%	5.1%	7.5%	9.7%	4.2%
ME	86	5.6%	2.9%	1.9%	3.4%	5.5%	7.3%	9.3%	3.9%
MD	214	5.5%	2.9%	2.2%	3.4%	5.2%	7.2%	9.0%	3.8%
CT	206	5.4%	2.9%	2.2%	3.4%	5.2%	6.9%	9.2%	3.5%
NH	71	5.4%	2.9%	2.1%	3.3%	5.3%	6.9%	9.2%	3.6%
MI	403	5.4%	3.0%	2.0%	3.3%	5.1%	7.0%	9.1%	3.7%
OH	896	5.4%	2.8%	2.1%	3.5%	5.1%	6.9%	8.9%	3.4%
NM	64	5.4%	3.0%	2.0%	3.0%	5.3%	7.1%	8.8%	4.1%
OR	109	5.3%	3.4%	1.5%	3.2%	4.8%	7.3%	9.7%	4.2%
WA	189	5.3%	3.0%	1.8%	2.9%	4.8%	6.9%	9.4%	4.0%
CO	194	5.2%	3.0%	1.7%	3.2%	5.0%	6.8%	9.2%	3.5%
LA	260	5.2%	3.0%	1.4%	3.3%	5.1%	6.9%	8.7%	3.7%
AZ	114	5.2%	3.2%	1.7%	3.0%	5.0%	6.4%	8.9%	3.4%
NJ	333	5.2%	2.7%	2.0%	3.4%	4.9%	6.7%	8.8%	3.2%
NV	52	5.2%	2.9%	1.7%	3.7%	4.7%	6.4%	9.1%	2.6%
WI	325	5.1%	2.9%	1.6%	3.1%	4.8%	6.9%	9.0%	3.8%
ND	78	5.1%	2.8%	1.6%	3.0%	4.8%	6.7%	9.2%	3.8%
OK	272	5.1%	3.6%	1.2%	2.5%	4.4%	7.0%	9.6%	4.5%
DC	16	5.0%	2.9%	1.7%	2.9%	4.8%	6.7%	9.7%	3.8%
AR	225	5.0%	3.1%	1.3%	2.6%	4.7%	7.0%	9.0%	4.3%
NE	187	4.9%	3.3%	1.5%	2.5%	4.5%	6.8%	9.3%	4.3%

State	Number of facilities	Mean score	Std dev.	10th percentile	25th percentile	50th percentile	75th percentile	90th percentile	Interquartile range
KS	300	4.9%	3.5%	1.0%	2.6%	4.7%	6.6%	8.8%	4.0%
FL	638	4.9%	2.7%	1.8%	3.1%	4.7%	6.4%	8.2%	3.3%
MO	489	4.9%	3.3%	1.2%	2.5%	4.3%	6.6%	9.3%	4.1%
ID	67	4.8%	3.0%	1.0%	2.3%	4.7%	6.8%	9.0%	4.5%
DE	44	4.6%	2.5%	2.0%	2.6%	4.2%	5.9%	8.5%	3.3%
MA	372	4.5%	2.4%	1.8%	2.9%	4.3%	6.0%	7.6%	3.0%
MN	334	4.5%	2.9%	1.5%	2.6%	3.9%	5.8%	8.4%	3.2%
HI	36	4.5%	1.7%	2.3%	2.9%	4.5%	5.6%	7.1%	2.7%
IA	406	4.3%	2.5%	1.3%	2.5%	4.1%	5.6%	7.8%	3.1%
TX	1126	4.3%	3.3%	0.7%	1.9%	3.7%	6.0%	8.5%	4.2%
RI	79	4.2%	2.0%	1.5%	2.9%	4.2%	5.4%	6.7%	2.6%
CA	1085	4.2%	2.7%	1.1%	2.3%	3.8%	5.6%	7.4%	3.3%
UT	78	4.1%	3.4%	0.8%	2.1%	3.4%	5.1%	10.0%	3.0%

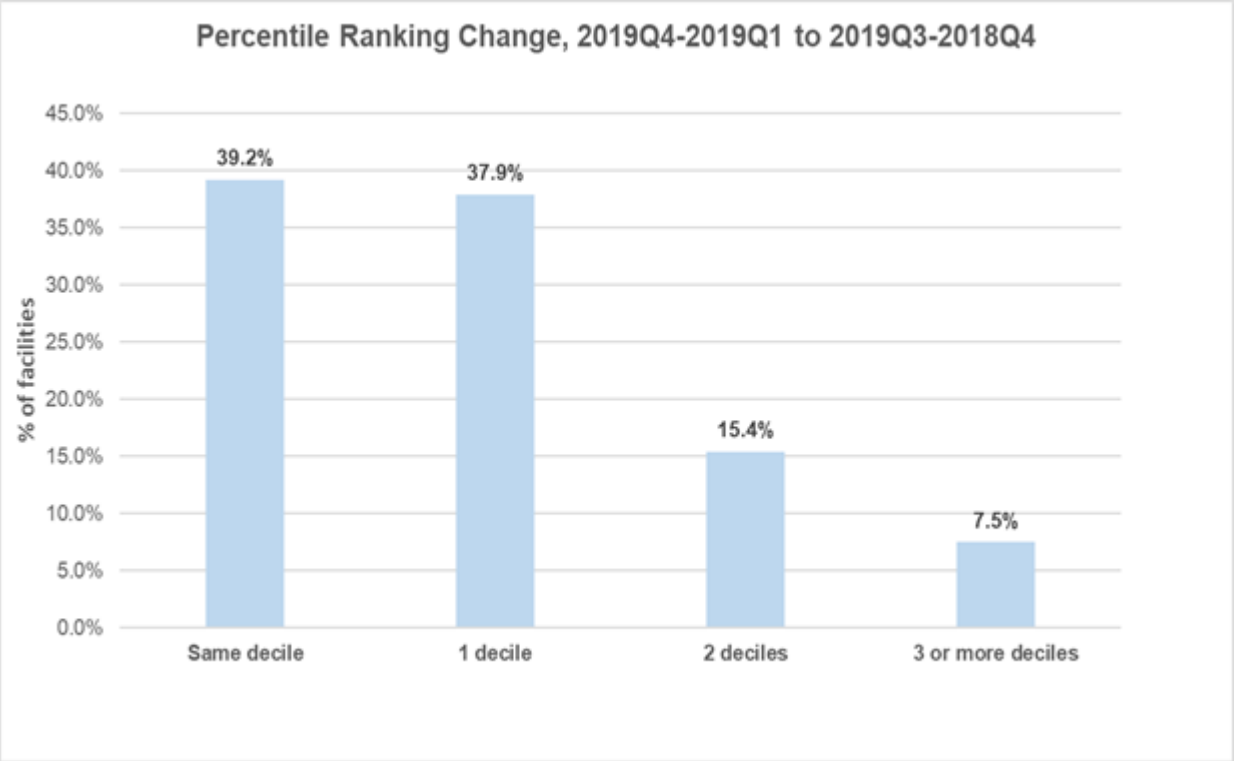
- Seasonality: This seasonal variation analysis was conducted at the quarterly level to identify any seasonal changes in measure scores. However, this measure is publicly reported on Care Compare based on rolling four quarters of data, which likely minimizes concerns about seasonality. For measure testing, the seasonal variation in the measure score was examined by plotting the mean and median national level scores for each quarter from 2018Q1- 2019Q4. The weight loss measure (NQF 0689) appears to be relatively stable over time with only slight seasonal variation being observed, as the mean and median scores for this measure saw spikes in 2019Q2. The results are presented in **Figure 1** below.

Figure 1. Trends over time for NQF #0689 Percent of Residents Who Lose Too Much Weight (Long Stay)



- Stability analysis: Figure 2 illustrates the changes in facility rank by quality measure score in two reporting periods by four rolling quarters: 2019Q4-2019Q1 and 2019Q3-2018Q4. Comparing ranks for these quarters, 39.2% of facilities' percentile rankings were constant within the same decile, 37.9% of facilities changed rank within 1 decile, 15.4% changed rank within 2 deciles, and 7.5% changed rank by 3 or more deciles.

Figure 2. Decile Change in QM Ranks between 2019Q4-2019Q1 and 2019Q3-2018Q4, NQF #0689 Percent of Residents Who Lose Too Much Weight (Long Stay)



- Confidence interval analysis: Table 4 shows the proportions of facilities that scored significantly higher or lower than the national facility-level mean. Data from 2019Q1-2019Q4 was combined to produce Table 4, as this range of data provides a better opportunity to capture low performing and high performing facilities. For this analysis, statistical significance was determined using 95% confidence intervals. A facility's quality measure score was statistically significantly different from the national mean if the national mean was not within that facility's 95% confidence interval. This analysis was also stratified by decile of facility size based on the number of residents who qualify for the denominator count.

For one year of data, 28.7% of facilities had a score that was statistically significantly different from the national mean with 95% confidence. Approximately 15.4% of facilities had scores that were statistically significantly lower than the national mean, and 13.2% of facilities had scores that were statistically significantly higher than the national mean with 95% confidence.

Table 4. Proportion of Facilities with Scores Significantly Different from the National Facility-Level Mean, Stratified by Facility Denominator Size for NQF #0679, FY2019 Q4

Decile of denominator size in residents	Facility Size (N Residents)	Facility Size (N Residents)	Mean QM Score	Number of facilities	Number of facilities with 95% confidence interval lower than national mean (%)	Number of facilities with 95% confidence interval lower than national mean (%)	Number of facilities with 95% confidence interval higher than national mean (%)	Number of facilities with 95% confidence interval higher than national mean (%)	Total number of facilities with scores significantly different from mean (%)	Total number of facilities with scores significantly different from mean (%)
*	Lower Bound	Upper Bound	*	*	N	%	N	%	N	%
1st Decile	20	108	5.08 %	1,409	193	13.7%	93	6.6%	286	20.3%
2nd Decile	109	137	5.06 %	1,422	140	9.8%	124	8.7%	264	18.6%
3rd Decile	138	165	5.04 %	1,413	203	14.4%	145	10.3%	348	24.6%
4th Decile	166	194	5.27 %	1,453	193	13.3%	174	12.0%	367	25.3%
5th Decile	195	223	5.13 %	1,434	216	15.1%	168	11.7%	384	26.8%
6th Decile	224	253	5.32 %	1,408	234	16.6%	210	14.9%	444	31.5%
7th Decile	254	288	5.19 %	1,448	269	18.6%	203	14.0%	472	32.6%
8th Decile	289	335	5.39 %	1,419	229	16.1%	237	16.7%	466	32.8%
9th Decile	336	422	5.43 %	1,428	225	15.8%	238	16.7%	463	32.4%
10th Decile	423	2,348	5.32 %	1,440	300	20.8%	296	20.6%	596	41.4%
Total	No data	No data	**5.22 %**	**14,274**	**2,202**	**15.4%**	**1,888**	**13.2%**	**4,090**	**28.7%**

* Indicates the table cell is empty

[Response Ends]

2b.04. Provide 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?)

[Response Begins]

[2021 submission] **Performance Measure Score Validity**

This measure is relatively stable over time with some limited seasonal variation being observed. Given the final public reporting form of this measure is a four quarter rolling average, seasonal variation is minimal in magnitude for this measure.

State-level variation is also a minimal source of variation. The proportion of variance in this measure explained by the state in which facilities are located was only 4.8% ($p < 0.001$).

The measure's correlations with related quality measures are all in the expected direction, which demonstrates strong convergence validity. This measure has modest negative correlations with Overall Facility Five-Star Ratings, Quality Ratings, Staffing Ratings, and Registered Nurse Staffing Ratings. This measure is modestly and positively correlated with Percent of Residents Whose Ability to Move Independently Worsened, Percent of Residents Whose Need for Help with Activities of Daily Living Has Increased, and Percent of Residents Who Have Depressive Symptoms. There may be several reasons for the relatively low correlation between NQF #0689 and these related measures, including the strong possibility that these measures simply capture different aspects of care. Additionally, some of these quality measures are also low frequency measures, which could contribute to the low correlations with Percent of Residents Who Lose Too Much Weight. While the majority of these correlation coefficients are modest in magnitude, they are all statistically significant and in the expected direction.

The confidence interval analysis for this measure indicates that there are meaningful differences in facility-level scores for this measure, as 28.7% of facilities had a mean score for which the 95% confidence intervals did not overlap with the national mean. The proportion of facilities with scores that were significantly different from the national mean varied as a function of the number of residents included in the denominator for this measure. In general, more facilities with a higher number of residents had scores significantly higher than the mean than those with a smaller number of residents. As the facility size increased, a greater proportion of facilities had scores that were significantly higher than the national mean (i.e. worse performers). However, a greater share of the largest sized facilities (10^{th} deciles) were observed to have performance significantly lower than the mean than the proportion of mid-sized facilities (i.e. those in the 5^{th} and 6^{th} deciles) and small-sized facilities (i.e. those in the 1^{st} and 2^{nd} deciles).

The stability analysis shows that while there were some changes from one quarter to another, 39.2% of facilities remained in the same decile and an additional 53.3% of facilities had score changes of 2 deciles or less in the next quarter.

[Response Ends]

2b.05. 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 in Importance to Measure and Report: Gap in Care/Disparities.

[Response Begins]

[2021 submission] In order to identify meaningful differences in facility performance on NQF #0689, the current variability in the facility-level quality measure scores was explored (see **2b.06**). The proportions of facilities with scores for this measure that are significantly different from the national facility-level mean were also explored and stratified by facility denominator size (see **2b.03**). For this analysis, statistical significance was determined using 95% confidence intervals: a facility's quality measure score was significantly different from the national mean if the national mean was not included in the facility's 95% confidence interval. High-performing facilities should have scores that are significantly above average, and scores of low-performing facilities should be significantly below average. The analysis was stratified by facility denominator size to examine whether this feature of the measure varies by size.

[Response Ends]

2b.06. Describe the statistical results from testing the ability to identify statistically significant and/or clinically/practically meaningful differences in performance measure scores across measured entities.

Examples may include 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.

[Response Begins]

[2021 submission] **Table 8** describes the current variability in the quality measure scores of facilities nationally, including both single quarter averages and the four-quarter average of quality measure scores in 2019. The four-quarter national mean facility-level score for this quality measure was 5.2% (2019Q1-2019Q4) with a median score of 4.8%. The interquartile range for this measure was 3.9 percentage points. Among facilities who were eligible to publicly report this measure, 2.6% (*k* = 372) had perfect scores of 0%.

Table 8. National Facility-Level Score Distribution, NQF #0689 Percent of Residents Who Lose Too Much Weight (Long Stay), 2019Q1-2019Q4

Quarter	*K*	Mean score	Std dev.	10th percentile	25th percentile	50th percentile	75th percentile	90th percentile	% of facilities with perfect score	Interquartile range
2019 Four Quarter Average	**14,274**	**5.2%**	**3.1%**	**1.6%**	**3.1%**	**4.8%**	**6.9%**	**9.2%**	**2.6%**	**3.9 % points**
2019Q4	13,976	5.3%	4.2%	0.0%	2.4%	4.6%	7.5%	10.8%	13.5%	5.1 % points
2019Q3	13,962	5.2%	4.1%	0.0%	2.3%	4.5%	7.3%	10.4%	14.3%	5.0 % points
2019Q2	13,977	5.4%	4.3%	0.0%	2.4%	4.8%	7.7%	11.1%	13.7%	5.3 % points
2019Q1	13,776	4.9%	4.1%	0.0%	2.0%	4.2%	7.1%	10.3%	16.8%	5.1 % points

Table 4 in Section **2b.03** shows the proportions of facilities that scored statistically significantly higher or lower than the national facility-level mean for one year of data (2019Q1-2019Q4). Overall, 28.7% of facilities scored significantly differently than the national mean between 2019Q1 and 2019Q4. The data were also stratified by the facility denominator size to examine the relationship between facility size and the reliability of facility scores. The proportions of facilities with scores that were significantly different from the national mean varied as a function of the number of residents included in the denominator for this measure.

[Response Ends]

2b.07. Provide 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.

In other words, what do the results mean in terms of statistical and meaningful differences?

[Response Begins]

[2021 submission] These analyses show that the quality measure score varies enough to make meaningful distinctions between high- and low-quality facilities. The 90th percentile is more than five times higher than the 10th percentile, and there is substantial distinction between the first and the third quintiles. Moreover, the quality measure scores vary sufficiently from the national mean that there are meaningful differences to differentiate the best and worst performers for this measure.

[Response Ends]

2b.08. Describe the method of testing conducted to identify the extent and distribution of missing data (or non-response) and demonstrate that performance results are not biased due to systematic missing data (or differences between responders and non-responders). Include how the specified handling of missing data minimizes bias.

Describe the steps—do not just name a method; what statistical analysis was used.

[Response Begins]
[2021 submission] Missing data represent a potential threat to the validity of a quality measure. Bias may be introduced if missing data is associated with resident or facility characteristics. Therefore, the rate of missing data per total number of assessments was examined. The results of this assessment are discussed in **Section 2b.09.**
[Response Ends]

2b.09. Provide the overall frequency of missing data, the distribution of missing data across providers, and the results from testing related to missing data.

For example, provide results of sensitivity analysis of the effect of various rules for missing data/non-response. If no empirical sensitivity analysis was conducted, identify the approaches for handling missing data that were considered and benefits and drawbacks of each).

[Response Begins]
[2021 submission] Based on analyses of MDS 3.0 data, missing data is not a threat to validity for this measure as very few resident episodes were excluded from the QM calculation due to missing data. An episode is excluded if:

1. The resident's target (latest qualifying) assessment is an OBRA Admission assessment (A0310A = [1]) ***or*** a PPS 5-Day assessment (A0310B = [1]), or
2. The resident's prognosis of life expectancy is less than 6 months (J1400 = [1]) or the Prognosis item is missing (J1400 = [-]) on the target assessment, or
3. The resident is Receiving Hospice care (O0100K2 = [1]) or the Hospice care item is missing (O0100K2 = [-]) on the target assessment, or
4. The weight loss item is missing (K0300= [-]) on the target assessment.

Only 1,551 episodes in the 2019 (Q1-Q4) long stay resident sample were excluded from the denominator for this measure due to missing responses on the prognosis of life expectancy being less than 6 months, which accounts for 0.04% of the total episodes. Additionally, only 7,948 (0.241%) episodes in the 2019 (Q1-Q4) long stay residents sample were excluded due to missing responses for the Hospice care item, and only 30,854 (0.935%) episodes were excluded due to missing responses for the weight loss item.

[Response Ends]

2b.10. Provide 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 non-responders), and how the specified handling of missing data minimizes bias.

In other words, 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 was conducted, justify the selected approach for missing data.

[Response Begins]
[2021 submission] There were too few residents excluded due to missing data to warrant concern over missing data introducing bias into the measure. Additionally, the number of excluded cases was too small to test for any kind of differences between facilities. Therefore, no further analyses were performed regarding missing data and this measure.
[Response Ends]

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 eCQMs). 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.

2b.11. Indicate whether there is more than one set of specifications for this measure.

[Response Begins]
No, there is only one set of specifications for this measure
[Response Ends]

2b.12. 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. Indicate what statistical analysis was used.

[Response Begins]
[Response Ends]

2b.13. Provide the statistical results from testing comparability of performance scores for the same entities when using different data sources/specifications.

Examples may include correlation, and/or rank order.

[Response Begins]
[Response Ends]

2b.14. Provide your interpretation of the results in terms of the differences in performance measure scores for the same entities across the different data sources/specifications.

In other words, what do the results mean and what are the norms for the test conducted.

[Response Begins]
[Response Ends]

2b.15. Indicate whether the measure uses exclusions.

[Response Begins]
Yes, the measure uses exclusions.
[Response Ends]

2b.16. Describe the method of testing exclusions and what was tested.

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?

[Response Begins]
[2021 submission] Please see **Section 2b.08**. "Missing data analysis and minimizing bias for analysis of this measure's exclusions," which are only for missing data on the applicable weight loss items.

[Response Ends]

2b.17. Provide 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.

[Response Begins]
[2021 submission] Please see **Section 2b.09**. "Missing data analysis and minimizing bias for analysis of this measure's exclusions," which are only for missing data on the applicable weight loss items.
[Response Ends]

2b.18. Provide your interpretation of the results, in terms of demonstrating that exclusions are needed to prevent unfair distortion of performance results.

In other words, 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.

[Response Begins]
[2021 submission] Please see **Section 2b.10**. "Missing data analysis and minimizing bias for analysis of this measure's exclusions," which are only for missing data on the applicable weight loss items.
[Response Ends]

2b.19. Check all methods used to address risk factors.

[Response Begins]
No risk adjustment or stratification
[Response Ends]

2b.20. If using statistical risk models, provide detailed risk model specifications, including the risk model method, risk factors, risk factor data sources, coefficients, equations, codes with descriptors, and definitions.

[Response Begins]
[2021 submission] Not applicable. This measure is not risk-adjusted.
[Response Ends]

2b.21. If an outcome or resource use measure is not risk-adjusted or stratified, provide rationale and analyses to demonstrate that controlling for differences in patient characteristics (i.e., case mix) is not needed to achieve fair comparisons across measured entities.

[Response Begins]
[2021 submission] The measure is not risk adjusted through a statistical model nor through stratification. However, clinical factors for risk adjustment were explored and a relevant clinical factor was tested statistically. The discussion is presented in **Sections 2b.23 - 2b.25** below. No ICD-10 diagnosis codes from item I8000 were considered for the risk adjustment.
[Response Ends]

2b.22. Select all applicable resources and methods used to develop the conceptual model of how social risk impacts this outcome.

[Response Begins]

Published literature
Internal data analysis
[Response Ends]

2b.23. Describe the conceptual and statistical methods and criteria used to test and select patient-level risk factors (e.g., clinical factors, social risk factors) used in the statistical risk model or for stratification by risk.

Please be sure to address the following: potential factors identified in the literature and/or expert panel; regression analysis; statistical significance of p<0.10 or other statistical tests; correlation of x or higher. Patient factors should be present at the start of care, if applicable. Also discuss any "ordering" of risk factor inclusion; note whether social risk factors are added after all clinical factors. Discuss any considerations regarding data sources (e.g., availability, specificity).

[Response Begins]
[2021 submission] **Risk Adjustor Selection – Conceptual Rationale and Statistical Testing**
Social Risk Factors

- Age: Older residents may be at higher risk for experiencing weight loss that is not due to a physician-prescribed weight loss regimen, as older age is associated with lower BMI (Tamura, Bell, Masaki, & Amella, 2013).
- Race: White residents may be at lower risk for experiencing weight loss that is not due to a physician-prescribed weight loss regimen than Black residents, as studies have shown that weight loss in long term care residents is associated with Black race/ethnicity (Newman, et al., 2002).

References:

1. Tamura, Bruce K., Bell, Christina L., Masaki, Kamal H., and Elaine J. Amella. (2013). Factors associated with weight loss, low BMI, and malnutrition among nursing home patients: a systematic review of the literature. Journal of the American Medical Directors Association, 14(9): 649-655. https://doi.org/10.1016/j.jamda.2013.02.022.
2. Newman, Anne B., Yanez, David, Harris, Tamara, Duxbury, Andrew, Enright, Paul L., and Linda P. Fried. (2002). Weight change in old age and its association with mortality. Journal of the American Geriatrics Society, 49(10): 1309–18. https://doi.org/10.1046/j.1532-5415.2001.49258.x.

[Response Ends]

2b.24. Detail the statistical results of the analyses used to test and select risk factors for inclusion in or exclusion from the risk model/stratification.

[Response Begins]
[2021 submission] **Internal data analysis**

Variables were created for the risk factor described above as follows:

- Age: Defined by the birth date reported on item A0900 on the MDS. A resident was defined as being aged if they were 85 years of age or older (as of the target assessment).
- Race: Defined by item A1000 on the MDS. A resident was defined as being white if item A1000F was checked in the MDS and non-white otherwise (as of the target assessment).

The results of the risk-adjustment model using age as a risk factor are summarized in **Table 5** below. Overall, the odds ratio for residents over the age of 85 is 1.17 and is statistically significant at the 5% level. The odds of experiencing weight loss is 1.17 times higher for residents over the age of 85 than for younger residents. However, the C-statistic (0.52) of the model indicates weak model performance and suggests that the model does not have high predictive ability.

Table 5. Assessment of Alternate Risk Adjustment Specifications: Age as the Covariate, NQF #0689 (2019Q4)

Model Covariates	Frequency of residents w/ covariate value	Frequency of residents with weight loss	% residents with weight loss	Odds Ratio	95% CI		C - statistic
Base case: Age < 85	591,289	29,886	5.1%	-	-	-	0.52
Age >= 85	340,080	19,995	5.9%	1.17	1.15	1.20	No data

Cells marked by a dash (-) are intentionally left blank.

The results of the risk-adjustment model using race as a risk factor are summarized in **Table 6** below. Overall, the odds ratio for White residents is 1.03 and is statistically significant at the 5% level. The odds of experiencing weight loss is 1.03 times higher for White residents than non-White residents. However, the C-statistic (0.50) of the model indicates weak model performance and suggests that the model does not have high predictive ability.

Table 6. Assessment of Alternate Risk Adjustment Specifications: Race as the Covariate, NQF #0689 (2019Q4)

Model Covariates	Frequency of residents w/ covariate value	Frequency of residents with weight loss	% of residents with weight loss	Odds Ratio	95% CI	-	C - statistic
Base case: Non-White Residents	259,408	13,573	5.2%	-	-	-	0.50
White Residents	671,961	36,308	5.4%	1.03	1.01	1.06	No data

Cells marked by a dash (-) are intentionally left blank.

The results of the risk-adjustment model using race and age as risk factors are summarized in **Table 7** below. Overall, the odds ratio for residents over the age of 85 is 1.17 and is statistically significant at the 5% level. The odds of experiencing weight loss is 1.01 times higher for White residents than non-White residents when controlling for age. However, this result is not statistically significant, indicating that White race may not be a strong risk factor for weight loss in this population. Additionally, the C-statistic (0.52) of the model indicates weak model performance and suggests that the model does not have high predictive ability.

Table 7. Assessment of Alternate Risk Adjustment Specifications: White Race and Age as the Covariates, NQF #0689 (2019Q4)

Model Covariates	Frequency of residents w/ covariate value	Frequency of residents with weight loss	% of residents with weight loss	Odds Ratio	P-value	95% CI	95% CI	C - statistic
Base case: Non-White Residents	181,545	9,572	5.3%	-	-	-	-	0.52
White Residents	671,961	36,308	5.4%	1.01	0.29	0.99	1.03	No data
Age >= 85	340,080	19,995	5.9%	1.17	<0.001	1.15	1.19	No data

Cells marked by a dash (-) are intentionally left blank.

[Response Ends]

2b.25. Describe the analyses and interpretation resulting in the decision to select or not select social risk factors.

Examples may include prevalence of the factor across measured entities, availability of the data source, empirical association with the outcome, contribution of unique variation in the outcome, or assessment of between-unit effects and within-unit effects. Also describe the impact of adjusting for risk (or making no adjustment) on providers at high or low extremes of risk.

[Response Begins]
[2021 submission] The analysis in **Section 2b.24** indicates that residents over the age of 85 are more likely to experience weight loss that is not due to a physician-prescribed weight loss regimen than younger residents. This observation is in the expected direction, as older age is typically associated with lower BMI in the nursing home population. The analysis in Section **2b.24** also indicates that White residents are more likely to experience weight loss that is not due to a physician-prescribed weight loss regimen than non-White residents. This observation is not in the expected direction, as research suggests that weight loss in long term care residents is associated with Black race/ethnicity (Newman, et al., 2002). Acumen hypothesizes that the higher average age of White residents compared to non-White residents may be driving in this unexpected observation. An additional risk adjustment model was run using both race and age as covariates. The results in Table 7 indicate that when controlling for age, the higher odds of experiencing weight loss for the White population than the non-White population is reduced (odds ratio moves from 1.03 to 1.01) and is no longer statistically significant. Therefore, White residents being more likely to experience weight loss that is not due to a physician-prescribed weight loss regimen than non-White residents is due to the higher average age of White residents. Although the results of the risk-adjustment models summarized in Tables 5 and 6 appear to be statistically significant at the 5% level, low C-Statistics were observed for both models. This suggests that the models do not have high predictive ability.
[Response Ends]

2b.26. 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). Provide the statistical results from testing the approach to control for differences in patient characteristics (i.e., case mix) below. If stratified ONLY, enter "N/A" for questions about the statistical risk model discrimination and calibration statistics.

Validation testing should be conducted in a data set that is separate from the one used to develop the model.

[Response Begins]
[2021 submission] This is not applicable. This measure is not risk-adjusted.
[Response Ends]

2b.27. Provide risk model discrimination statistics.

For example, provide c-statistics or R-squared values.

[Response Begins]
[2021 submission] This is not applicable. This measure is not risk-adjusted.
[Response Ends]

2b.28. Provide the statistical risk model calibration statistics (e.g., Hosmer-Lemeshow statistic).

[Response Begins]
[2021 submission] This is not applicable. This measure is not risk-adjusted.
[Response Ends]

2b.29. Provide the risk decile plots or calibration curves used in calibrating the statistical risk model.

The preferred file format is .png, but most image formats are acceptable.

[Response Begins]

[2021 submission] This is not applicable. This measure is not risk-adjusted.
[Response Ends]

2b.30. Provide the results of the risk stratification analysis.

[Response Begins]
[2021 submission] This is not applicable. This measure is not risk-adjusted.
[Response Ends]

2b.31. Provide your interpretation of the results, in terms of demonstrating adequacy of controlling for differences in patient characteristics (i.e., case mix).

In other words, what do the results mean and what are the norms for the test conducted?

[Response Begins]
[2021 submission] This is not applicable. This measure is not risk-adjusted.
[Response Ends]

2b.32. Describe any additional testing conducted to justify the risk adjustment approach used in specifying the measure.

Not required but would provide additional support of adequacy of the risk model, e.g., testing of risk model in another data set; sensitivity analysis for missing data; other methods that were assessed.

[Response Begins]
[2021 submission] This is not applicable. This measure is not risk-adjusted.
[Response Ends]

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.

3.01. Check all methods below that are used to generate the data elements needed to compute the measure score.

[Response Begins]
Generated or collected by and used by healthcare personnel during the provision of care (e.g., blood pressure, lab value, diagnosis, depression score)
[Response Ends]

3.02. Detail to what extent the specified data elements are available electronically in defined fields.

In other words, indicate whether data elements that are needed to compute the performance measure score are in defined, computer-readable fields.
[Response Begins]
ALL data elements are in defined fields in electronic clinical data (e.g., clinical registry, nursing home MDS, home health OASIS)
[Response Ends]

3.03. 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 data elements not from electronic sources.

[Response Begins]
This is not applicable; all data elements used to calculate the measure are in defined fields in electronic clinical data.
[Response Ends]

3.04. Describe any efforts to develop an eCQM.

[Response Begins]
There are no current efforts to develop this measure as an eMeasure.
[Response Ends]

3.06. 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.

[Response Begins]
The general data collection method for the MDS 3.0 is currently in operational use and mandatory for all Medicare/Medicaid certified nursing facilities. Missing data is not a threat to the validity for this measure, as only 1,551 (0.04%) of episodes in 2019 were excluded from the denominator due to missing responses in prognosis of life expectancy being less than 6 months. Additionally, only 7,948 (0.241%) of episodes in 2019 were excluded due to missing response for the Hospice care item, and only 30,854 (0.935%) of episodes were excluded due to missing response for the weight loss item. The missing data is too small to warrant concern over introducing bias into the measure.
[Response Ends]

Consider implications for both individuals providing data (patients, service recipients, respondents) and those whose performance is being measured.

3.07. Detail 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),

Attach the fee schedule here, if applicable.

[Response Begins]
This is not applicable.
[Response Ends]

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.

Extent to which intended audiences (e.g., consumers, purchasers, providers, policy makers) can understand the results of the measure and are likely to find them useful for decision making.

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 demonstrating performance improvement.

4a.01.

Check all current uses. For each current use checked, please provide:

Name of program and sponsor

URL

Purpose

Geographic area and number and percentage of accountable entities and patients included

Level of measurement and setting

[Response Begins]
Public Reporting
Quality Improvement with Benchmarking (external benchmarking to multiple organizations)
Quality Improvement (Internal to the specific organization)
Public Reporting:

- Program and sponsor: Care Compare and Provider Data Catalog/Centers for Medicare and Medicaid (CMS)
- URL:
    - Care Compare: https://www.medicare.gov/care-compare/
    - Provider Data Catalog: https://data.cms.gov/provider-data/
- Purpose: Consumer information
- Geographic area and number and percentage of accountable entities and patients included: All United States Nursing Homes with Medicare-eligible long-stay residents. Between 2019Q1 and 2019Q4 there were 15,332 eligible facilities and 1,247,968 residents with target assessments, and 14,274 facilities (93.1%) had sufficient sample size (20 or more long-stay residents included in the denominator) to report on this measure, and 1,230,998 residents (98.6%) were included in the calculation of this measure. Four individual quarter scores are publicly reported on Provider Data Catalog. To enhance measurement stability and reliability beyond a one-quarter measure, a four-quarter average version of the measure is publicly reported as part of the Nursing Home Quality Initiative (NHQI) through Care Compare and Provider Data Catalog.

Quality Improvement with Benchmarking (external benchmarking to multiple organizations):

- Program and sponsor: Certification and Survey Provider Enhanced Reports (CASPER)/Centers for Medicare and Medicaid
- URL: https://qtso.cms.gov/
- Purpose: Quality improvement
- Geographic area and number and percentage of accountable entities and patients included: All United States Medicare/Medicaid certified Nursing Homes with eligible long-stay residents regardless of denominator sample

size. Between 2019Q1 and 2019Q4 there were 15,332 eligible facilities and 1,247,968 residents with target assessments.

Quality Improvement (internal to the specific organization):

- Program and sponsor: Certification and Survey Provider Enhanced Reports (CASPER)/Centers for Medicare and Medicaid
- URL: https://qtso.cms.gov/
- Purpose: Quality improvement
- Geographic area and number and percentage of accountable entities and patients included: All United States Medicare/Medicaid certified Nursing Homes with eligible long-stay residents regardless of denominator sample size. Between 2019Q1 and 2019Q4 there were 15,332 eligible facilities and 1,247,968 residents with target assessments.

[Response Ends]

4a.02. Check all planned uses.

[Response Begins]
Public reporting
Quality Improvement with Benchmarking (external benchmarking to multiple organizations)
Quality Improvement (internal to the specific organization)
[Response Ends]

4a.03. If not currently publicly reported OR used in at least one other accountability application (e.g., payment program, certification, licensing), explain why the measure is not in use.

For example, do policies or actions of the developer/steward or accountable entities restrict access to performance results or block implementation?

[Response Begins]
This is not applicable, this measure is publicly reported.
[Response Ends]

4a.04. If not currently publicly reported OR used in at least one other accountability application, provide a credible plan for implementation within the expected timeframes: used in any accountability application within 3 years, and publicly reported within 6 years of initial endorsement.

A 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.

[Response Begins]
This is not applicable, this measure is publicly reported.
[Response Ends]

4a.05. Describe how performance results, data, and assistance with interpretation have been provided to those being measured or other users during development or implementation.

Detail 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.

[Response Begins]

This quality measure (NQF 0689, Percent of Residents Who Lose Too Much Weight (Long Stay)) is part of the Nursing Home Quality Initiative (NHQI). Information on this measure is available to both nursing home providers and to the public.

All United States Medicare and/or Medicaid certified nursing home providers may view their performance results for this and other NHQI measures via the Certification and Survey Provider Enhanced Reports (CASPER) system. These CASPER MDS 3.0 QM reports are intended to provide nursing home providers with feedback on their quality measure scores, helping them to improve the quality of care delivered to their residents. CASPER MDS 3.0 reports also include Resident-Level Quality Measure Reports, which allow providers to identify the residents that trigger a particular quality measure (by scanning a column of interest and looking for the residents with an "X") and to identify residents who trigger multiple quality measures. Providers can use this information to target residents for quality improvement activities. Quality measure reports are also available to state surveyors and facility staff through the CASPER reporting system.

Consumers, including current and prospective nursing home residents and their families/caregivers, may access nursing home performance scores on this quality measure via the Care Compare website (https://www.medicare.gov/care-compare/?providerType=NursingHome) or the Provider Data Catalog (https://data.cms.gov/provider-data/). The Care Compare site reports the four-quarter average, while the Provider Data Catalog site reports the one-quarter version of the measure alongside the four-quarter average.

Further, providers have an opportunity to review their performance prior to public reporting on the Care Compare website via Provider Preview Reports, also available through the CASPER system. These reports allow providers to view their quality measure scores for each NHQI measure, along with state and national averages for comparison, to identify potential errors in data submission or other information and request an update. Detailed instructions on how to view and interpret reports, including an explanation of differences between the quality measure reports and publicly reported information, are provided in the CASPER Reporting MDS Provider Users Guide, Section 11, which can be found at the following website: https://qtso.cms.gov/system/files/qtso/cspr_sec11_mds_prvdr_0.pdf
[Response Ends]

4a.06. Describe the process for providing measure results, including when/how often results were provided, what data were provided, what educational/explanatory efforts were made, etc.

[Response Begins]
The CASPER reports are available to providers on-demand with quality measure data updated monthly. Care Compare reports the rolling average of four quarters for the quality measure, comparing each nursing home's score to both the state and national average; providers can preview this information before it is publicly reported.

Detailed instructions on how to view and interpret reports, including an explanation of differences between the quality measure reports and publicly reported information, are provided in the CASPER Reporting MDS Provider Users Guide, Section 11, at the following website: https://qtso.cms.gov/system/files/qtso/cspr_sec11_mds_prvdr_0.pdf

CMS provides a Help Line, which is accessible by telephone and email, to answer provider questions about the NHQI quality measures and reporting requirements.
[Response Ends]

4a.07. Summarize the feedback on measure performance and implementation from the measured entities and others. Describe how feedback was obtained.

[Response Begins]
CMS is committed to receiving ongoing feedback on measures implemented as part of the NHQI. CMS takes into consideration feedback and input on measure performance and implementation through the appropriate sub-regulatory communication channels, including but not limited to: NQF public comment periods held as part of endorsement processes; feedback from providers submitted to the CMS quality measure support inboxes and feedback from the provider community on Open Door Forums (ODFs). CMS has not received any feedback on measure performance or implementation.
[Response Ends]

4a.08. Summarize the feedback obtained from those being measured.

[Response Begins]
Upon review of all inquiries submitted to the quality measure support inbox between 10/2019 and 09/2021, those being measured raised no concerns regarding the performance and implementation of NQF 0689.
[Response Ends]

4a.09. Summarize the feedback obtained from other users.

[Response Begins]
Upon review of all inquiries submitted to the quality measure support inbox between 10/2019 and 09/2021, other users raised no concerns regarding the performance and implementation of NQF 0689.
[Response Ends]

4a.10. Describe how the feedback described has been considered when developing or revising the measure specifications or implementation, including whether the measure was modified and why or why not.

[Response Begins]
This is not applicable.
[Response Ends]

4b.01. You may refer to data provided in Importance to Measure and Report: Gap in Care/Disparities, 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, provide an explanation. 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.

[Response Begins]
Progress (trends in performance results, number and percentage of people receiving high-quality healthcare)
The national facility-level mean and median scores for the Percent of Residents Who Lose Too Much Weight demonstrate stability from quarter to quarter (Figure 1 in the "Scientific Acceptability: Validity – Testing" section of this form). Overall, the national facility-level mean and median scores have remained relatively stable throughout the study period with only a slight increase in 2019Q2. The mean score for this measure was 5.24% in 2018Q1 and the median score was 4.55%. In 2019Q2, the mean and median were 5.44% and 4.76%, respectively. However, when facility-level mean and median scores are examined over an extended period of time, the weight loss measure exhibits some improvement. In 2011Q1, the mean score for this measure was 6.4% and the median score was 5.8%. In 2019Q1, the mean score for this measure was 4.9% and the median score was 4.17%. This decrease in facility-level mean and median scores between 2011Q1 and 2019Q1 suggests that the percent of residents who lose too much weight has decreased at the facility level over time.

Geographic area and number and percentages of accountable entities and patients included:
All United States Nursing Homes with Medicare-eligible long-stay residents. Between 2019Q1 and 2019Q4 there were 15,332 eligible facilities and 1,247,968 residents with target assessments, and 14,274 facilities (93.1%) had sufficient sample size (20 or more long-stay residents included in the denominator) to report on this measure, and 1,230,998 residents (98.6%) were included in the calculation of this measure.

[Response Ends]

4b.02. Explain any unexpected findings (positive or negative) during implementation of this measure, including unintended impacts on patients.

[Response Begins]
During the testing process for NQF 0689, the results of the risk-adjustment model using race as a risk factor demonstrated that white residents are more likely to experience weight loss that is not due to a physician-prescribed weight loss regimen than non-white residents (see the "Scientific Acceptability: Validity - Other Threats to Validity (Exclusions, Risk Adjustment)" section of this form). This observation was not in the expected direction, as research suggests that weight loss in long term care residents is associated with Black race/ethnicity (Newman, et al., 2002). Therefore, an additional risk adjustment model was run using both race and age as covariates. The results indicate that when controlling for age, the higher odds of experiencing weight loss for the White population than the non-White population is reduced (odds ratio moves from 1.03 to 1.01) and is no longer statistically significant. Therefore, white residents being more likely to experience weight loss that is not due to a physician-prescribed weight loss regimen than non-white residents is due to the higher average age of white residents.
[Response Ends]

4b.03. Explain any unexpected benefits realized from implementation of this measure.

[Response Begins]
This is not applicable; there are no unexpected benefits from the implementation of NQF 0689.
[Response Ends]

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.

If you are updating a maintenance measure submission for the first time in MIMS, please note that the previous related and competing data appearing in question 5.03 may need to be entered in to 5.01 and 5.02, if the measures are NQF endorsed. Please review and update questions 5.01, 5.02, and 5.03 accordingly.

5.01. Search and select all NQF-endorsed related measures (conceptually, either same measure focus or target population).

(Can search and select measures.)
[Response Begins]
[Response Ends]

5.02. Search and select all NQF-endorsed competing measures (conceptually, the measures have both the same measure focus or target population).

(Can search and select measures.)
[Response Begins]
[Response Ends]

5.03. If there are related or competing measures to this measure, but they are not NQF-endorsed, please indicate the measure title and steward.

[Response Begins]
There are no related or competing measures that are not NQF-endorsed.
[Response Ends]

5.04. If this measure conceptually addresses EITHER the same measure focus OR the same target population as NQF-endorsed measure(s), indicate whether the measure specifications are harmonized to the extent possible.

[Response Begins]
No
[Response Ends]

5.05. If the measure specifications are not completely harmonized, identify the differences, rationale, and impact on interpretability and data collection burden.

[Response Begins]
This is not applicable, as there are no related measures.
[Response Ends]

5.06. Describe why this measure is superior to competing measures (e.g., a more valid or efficient way to measure quality). Alternatively, justify endorsing an additional measure.

Provide analyses when possible.

[Response Begins]
This is not applicable, as there are no competing measures.
[Response Ends]

Appendix

Supplemental materials may be provided in an appendix.: Available in attached file
Attachment: 0689_NQF_0689_Measure_Submission_Appendix_20211101.docx

Contact Information

Measure Steward (Intellectual Property Owner) : Centers for Medicare & Medicaid Services
Measure Steward Point of Contact: Natanov, Rebekah, rebekah.natanov@cms.hhs.gov
Measure Developer if different from Measure Steward: Centers for Medicare & Medicaid Services
Measure Developer Point(s) of Contact: Santhosh, Aathira, asanthosh@sphereinstitute.org

Additional Information

1. Provide any supplemental materials, if needed, as an appendix. All supplemental materials (such as data collection instrument or methodology reports) should be collated one file with a table of contents or bookmarks. If material pertains to a specific criterion, that should be indicated.

[Response Begins]
Available in attached file
[Response Ends]

Attachment: 0689_NQF_0689_Measure_Submission_Appendix_20211101.docx

Appendix A: Convergent Validity Testing
Acumen's submission materials contain both data element validity testing and empirical validity testing. The testing for data element validity strongly demonstrates that the MDS assessment data elements used in the measure accurately capture the occurrence of unintended weight loss. Correlations in item responses between two gold-standard nurses and between a gold-standard and facility nurse are 0.944 and 0.918, respectively, showing that the measure relies on items that validly reflect what the measure intends to assess. Acumen also conducted empirical validity testing, demonstrating the ability to distinguish a substantial portion of facilities from the national mean, stability across time, the absence of any unexpected/concerning trends by season or state, and convergent validity. This appendix provides an additional discussion of convergent validity, in response to the Scientific Methods Panel (SMP) review. This discussion helps clarify confusion arising from the fact that several SMP members reviewed the incorrect testing results from the 2015 measure submission materials, rather than the materials Acumen submitted in this evaluation round.

During the SMP meeting, the lead discussant mentioned that convergent validity should demonstrate that better processes lead to better outcomes. Acumen agrees that providers that follow better processes should have lower rates of unintended weight loss. Our literature research shows that processes to lower unintended weight loss include:

- Early detection of weight loss
- Better eating environment
- Provision of nutritional supplements
- Long staff tenure and training
- Presence of dietician in the facility
- More certified nursing assistant and non-nursing staff hours per resident
- Positive relationship and better communication between staff and resident

Due to the nature of MDS assessments, facility-level data, and resident billing data, these processes are not directly observable in data. Therefore, Acumen's approach is to correlate unintended weight loss with measures that are impacted by similar processes. For example, a positive relationship between staff and a resident can contribute to a lower rate of depression; provision of nutritional supplements like evening snacks can improve functional status among mal-nourished residents. Facilities with a good process to promptly identify negative events can better identify and treat injuries from falls, which can in turn contribute to functional improvement. Such facilities would be better at, for example, early detection of stage 1 pressure ulcers and successfully reduce the chance of more severe pressure ulcers.

The above mentioned rationale sets our expectation that unintended weight loss should have a positive correlation with depression, function and pressure ulcer measures, because these measures share some similarities in good processes. At the same time, these correlations are expected to be moderately low (e.g. 0.1-0.2) – despite sharing some underlying processes, these measures assess distinct clinical outcomes that are shaped by many distinct processes, and some measures even have different denominator definitions. For example, the two function measures exclude residents who were previously totally dependent, which may contribute to a low correlation with the weight loss measure. Moreover, while literature suggests that low function is associated with weight loss, Acumen expects low correlation between weight loss and the two long-stay function measures, as both measures evaluate the change in functional status rather than the level of functional dependency. Change in functional status could be a result of orthopedic reasons unrelated to weight loss. Such expectations for the size of correlations across quality measures are consistent with CMS design, since measures on Nursing Home Compare should assess different aspects of quality and not contain redundant or duplicative measures.

Our results confirm these expectations and support the validity of measure 0689. Results in Section 2b.03 of this submission confirm the expected patterns for depression and function. Additionally, Acumen investigated convergent validity using the pressure ulcer measure – Percent of High-Risk Residents with Pressure Ulcers (0679) – and observed a statistically significant positive correlation of 0.15 between the weight loss measure and the pressure ulcer measure, which is again in the expected direction.

Although the convergent validity testing for the weight loss measure resulted in modest correlations in the hypothesized direction, some SMP panelists expressed concerns about the magnitude of the correlations, with emphasis on low correlations between the weight loss measure and global measures such as facility star ratings (e.g. for staffing). As noted just above, correlations with other quality measures were expected to be modest; moreover, the magnitudes are consistent with correlations observed in convergent validity testing for other NQF endorsed measures (e.g. Percent of Residents Experiencing One or More Falls with Major Injury (0674) and Percent of Residents with a Urinary Tract Infection (0684)). For correlations with staffing star ratings, literature demonstrates that certified nursing assistant (CNA) and non-nursing staff can both play a key part in determining the prevalence of weight loss in a facility. However, the five-star staffing measures rate facilities by registered nurse (RN) hours and total nursing hours. The difference in the type of staff may contribute to lower than expected correlations with the weight loss measure.

Appendix B: Risk Adjustment

Measure 0689 is intentionally not risk adjusted. Several SMP panelists expressed concern in their preliminary analysis that the need for risk adjustment was not adequately explored, and that particular comorbidities were not included in the risk adjustment analysis. Therefore, Acumen estimated risk adjustment models using the risk factors suggested by the panelists – and a series of other potential risk factors – to address these concerns. The results demonstrate that, while items suggested by SMP panelists generally have the expected correlation with the weight loss outcome, impacts on measure scores are small and the most predictive covariates are items that are under the control of the facility and hence inappropriate for inclusion in risk adjustment. Given this, the results support the decision to not risk adjust the measure. CMS and Acumen intend to continue investigating the potential for risk adjusting the measure in coming years. We briefly summarize the testing results below.

First, Acumen ran a logistic regression model with Alzheimer's, dementia, and depression as covariates, as shown in *Table 1*. The c-statistic is 0.51. The c-statistic represents how well the model is able to distinguish between positive and negative outcomes. A c-statistic of 0.51 suggests that the model is only slightly better in classifying outcomes than a coin toss (c-statistic of 0.5). Using this model, Acumen also constructed a risk-adjusted version of the measure. Acumen observed that the mean, median, standard deviation and interquartile range for the risk-adjusted weight loss measure are extremely close to the observed mean, median, standard deviation and interquartile range. When comparing facility ranking between the risk-adjusted score and observed score, Acumen observed that risk adjustment has very limited impact on provider ranking. Approximately 97% of facilities' percentile ranking was constant within the same decile, 3% changed rank within 1 decile, and no facilities changed rank by 2 or more deciles. This analysis suggests that while the covariates could have small effects on outcomes on a resident-level basis, they have minimal impact on the provider-level measure scores. Importantly, as discussed in the SMP meeting, the small effects on outcomes observed in Table 1 could derive from the fact that facilities pay special attention to weight loss in residents with these characteristics. Risk-adjusting for these characteristics could undermine this incentive for maintaining an important patient safety outcome for all residents, and so should be approached with caution. Fortunately, the results demonstrate that the measure is not meaningfully affected by exercising the appropriate caution to not risk adjust.

Table 1: Logistic Regression Using Alzheimer's, Dementia, and Depression

Model Covariates	**% of residents w/ weight loss**	**Odds Ratio**	**P-value**	**C - statistic**
Alzheimer's	5.7%	1.08	<.0001	0.51
Dementia	5.5%	1.06	<.0001	-
Depression	5.3%	0.96	<.0001	-

Cells marked by a dash (-) are intentionally left blank.

In addition to the three conditions above, the SMP also recommended considering age, cancer, swallowing disorder and diet related items for risk adjustment. Acumen has some concerns about risk adjusting for these items:

- **Age:** Acumen tested age as a risk factor for weight loss and observed that residents over the age of 85 are at higher risk for weight loss compared to residents under the age of 85 in Section 2b.24 of the submission form. However, risk adjustment by age can mask disparities in care across nursing homes with distinct age compositions. CMS has long-established guidance to ensure disparities in care associated with certain social risk factors are transparent to the public. Traditional risk adjustment models can have the unintended consequence of justifying insufficient care provided to patients at risk of patient safety outcomes, which may in turn result in disparities in patient safety.
- **Cancer:** Several panelists recommended including cancer as a potentially addressable risk factor for unintended weight loss. However, the MDS active diagnosis item I0100 for cancer was added to the quarterly MDS assessment in October 2019. Since the weight loss measure is reported as a four-quarter measure on Care Compare, Acumen did not have enough consistent data collection for this MDS item during our study window (2019Q1-2019Q4). A more recent study window beyond October 2019 was not used for testing the cancer item due to the ongoing data quality concerns caused by the COVID-19 pandemic. Acumen is open to consider risk adjusting the weight loss measure by cancer in the future when more data becomes available for testing.
- **Food Consumption:** One panelist recommended risk adjusting by the proportion of food left uneaten by a resident. However, the MDS does not contain information on the amount of food a resident eats or leaves. The only food intake item on the MDS is K0710B *Average fluid intake per day by IV or tube feeding,* which is only recorded for a small number of residents who need IV or tube feeding.
- **Diet and Swallowing Disorder:** One panelist recommend risk adjusting by pureed food (mechanically altered diet) and swallowing disorder. Acumen recommends delaying any risk adjustment by these two items due to data element reliability and validity concerns. First, mechanically altered diet is a service provision item under the control of the facility, which may be inappropriate to include as a risk factor. Second, these two items became payment items with the implementation of the new SNF PPS Patient Driven Payment Model (PDPM) that took effect October 1, 2019. Since the implementation of the new payment system, coding of mechanically altered diet and swallowing disorder among the SNF population has increased. For example, coding of swallowing disorder has surged from 4% to over 17%. Effective 10/1/2020, the Nursing Home Comprehensive (NC) and Nursing Home Quarterly (NQ) OBRA assessments can be used by states to collect PDPM HIPPS codes if states decide they want to participate in this data collection. Over 30 states have opted-in and may switch to a PDPM-based Medicaid payment in the near future. Based on coding changes observed among the SNF population, it is possible there will be similar changes for the nursing home population. It will be concerning if coding patterns of these two items differ across states due to difference in payment incentives rather than patient composition. Acumen recommends reevaluating the use of these two items in risk adjustment after some states transition to PDPM.

Despite these conceptual concerns, Acumen estimated a risk adjustment model using the above mentioned items (except cancer). The results displayed in Table 2 are useful to ensure that estimated relationships are as expected and that the impacts on measure scores are limited in any case. This risk adjustment model has a moderate c-statistic of 0.60, suggesting that the model has moderate but limited predictive power. Even with this more complicated model where the potentially problematic covariates are included, the change in provider ranking is limited: approximately 78% of facilities' percentile ranking was constant within the same decile, 21.4% changed rank within 1 decile, 0.5% changed rank within 2 deciles, and 0.1% changed rank by 3 or more deciles. Given the above-mentioned concerns and the limited impact on provider ranking, Acumen believes it is premature to include the above mentioned items in risk adjustment.

Table 2: Logistic Regression, Adding MAD, Swallowing Disorder, and Age

Model Covariates	% of residents w/ weight loss	Odds Ratio	P-value	C - statistic
Mechanically altered diet	7.2%	1.61	<.0001	0.60
No swallowing disorder	4.8%	0.47	<.0001	-
Age >= 85	5.9%	1.13	<.0001	-
Alzheimer's	5.7%	1.04	<.0001	-
Dementia	5.5%	1.05	<.0001	-
Depression	5.3%	1.08	<.0001	-

Cells marked by a dash (-) are intentionally left blank.

One panelist expressed interest in seeing the effect of a large range of comorbidities on weight loss, without following the more conventional practice for these measures of first selecting comorbidities based on literature search or clinical expectation (a practice which led to the comorbidities in Tables 1 and 2 above). Acumen would like to note that while using a wide range of comorbidities is common for claims-based post-acute care (PAC) quality measures, it is likely not a good practice for nursing home assessment-based measures. For those PAC claims-based measures, the broad sets of comorbidities are generally collected from the prior proximal hospital stays or from Part A/B claims in a 90-day or year-long lookback period. This helps ensure that conditions are present upon entry into the PAC stay. However, nursing home long stay residents reside in facilities for long durations of time. The median length of nursing home long stay is over two years long. Depending on the specific condition, comorbidities recorded on the assessment can be a reflection of the nursing home's care quality rather than conditions the residents come to the facilities with. Examples of such conditions are fracture as a result of falls, pressure ulcer, urinary tract infection, etc.
Despite these conceptual concerns, Acumen estimated a model using all MDS Section I Active Diagnoses items and Section L Oral/Dental Status items with the exception of a handful of diagnoses not collected on quarterly assessments in the study period or that are already a quality measure (Percent of Residents with a Urinary Tract Infection (0684)). It is important to note that Section I of the MDS consists of checkboxes for each listed active diagnosis and does not record ICD-10 codes. The inclusion of Section I and Section L items in the risk adjustment model expands the covariate list to 44 comorbidities. The c-statistic of this model is 0.57, which is quite low considering the number of covariates included. The AIC and BIC statistics of this model are very close to those of the 3-regressor model in *Table 1*, which suggests minimal improvement in model fit after adjustment for model complexity. Acumen also compared provider performance before and after risk adjustment using this model, and observed only small impacts on provider performance: 85.8% of facilities' percentile ranking was constant within the same decile, 14.2% changed rank within 1 decile, and less than 0.05% facilities changed rank by 2 or more deciles.
Based on the above mentioned conceptual considerations and risk adjustment results, Acumen believes the current approach of not risk adjusting the weight loss measure is appropriate. The necessity and appropriateness of risk adjusting this measure should be further evaluated in the future, with focus on the cancer item added in late 2019 and the payment incentives associated with items such as mechanically altered diet and swallowing disorder.

2. List the workgroup/panel members' names and organizations.

Describe the members' role in measure development.

[Response Begins]
Barbara Anglin, RN
Program Services Consultant
American Association of Nurse Assessment Coordinators (AANAC)

Bonnie Burak-Danielson, MSM, EXP, LPTA
Rehab Manager of Reimbursement
Spaulding Rehab Network

Sarah Burger, MPH, RN
Senior Advisor and Coordinator
Coalition of Geriatric Nursing Organizations
The John A. Hartford Institute for Geriatric Nursing

Diane Carter, MSN, RN, CS
President
AANAC

Kate Dennison, RN, RAC-MT
Minimum Data Set (MDS) Coordinator
The Cedars

Mary Ellard, RN, MPA/H, RAC-CT

Clinical Assessment Specialist
Five Star Quality Care, Inc.

Sandy Fitzler, RN
Senior Director of Clinical Services
American Health Care Association

David F. Hittle, PhD
Assistant Professor
Division of Health Care Policy and Research
University of Colorado Denver, School of Medicine

Steve Levenson, MD, CMD
Multi-Facility Medical Director, Baltimore, MD

Carol Maher, RN-BC, RAC-CT
Director of Clinical Reimbursement
Ensign Facilities Services

Barbara Manard, PhD
Vice President, Long Term Care/Health Strategies
American Association of Homes and Services for the Aging

Debra Saliba, MD, MPH
Anna and Harry Borun Chair in Geriatrics and Gerontology at UCLA
Research Physician VA GLAHS GRECC
Director of UCLA/JHA Borun Center for Gerentological Research
Senior Natural Scientist RAND Health
University of California, Los Angeles (UCLA), Veterans Affairs (VA), RAND Corporation

Eric Tangalos, MD
Professor of Medicine
Mayo Clinic

Jacqueline Vance, RNC, CDONA/LTC
Director of Clinical Affairs
American Medical Directors Association

Mary Van de Kamp, MS/CCC-SLP
Vice President, Clinical Rehabilitation
Peoplefirst Rehabilitation

Charlene Harrington, PhD, RN, FAAN*
Professor Emeritus University of California, San Francisco
Fellow in the American Academy of Nursing

This technical expert panel met during 2 days in January 2009 to review an environment scan of the current quality measures and make recommendations regarding their transition from MDS 2.0 to MDS 3.0.
[Response Ends]

3. Indicate the year the measure was first released.

[Response Begins]
2002
[Response Ends]

4. Indicate the month and year of the most recent revision.

[Response Begins]
04/2015
[Response Ends]

5. Indicate the frequency of review, or an update schedule, for this measure.

[Response Begins]
Every 3 years.
[Response Ends]

6. Indicate the next scheduled update or review of this measure.

[Response Begins]
04/2021
[Response Ends]

7. Provide a copyright statement, if applicable. Otherwise, indicate "N/A".

[Response Begins]
N/A
[Response Ends]

8. State any disclaimers, if applicable. Otherwise, indicate "N/A".

[Response Begins]
N/A
[Response Ends]

9. Provide any additional information or comments, if applicable. Otherwise, indicate "N/A".

[Response Begins]
N/A
[Response Ends]