NATIONAL QUALITY FORUM

MEASURE WORKSHEET

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

To navigate the links in the worksheet: Ctrl + click link to go to the link; ALT + LEFT ARROW to return

Brief Measure Information

NQF #: 0686

Corresponding Measures:

De.2. Measure Title: Percent of Residents Who Have/Had a Catheter Inserted and Left in Their Bladder (Long Stay)

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

De.3. Brief Description of Measure: This measure reports the percentage of low risk, long-stay residents who have had an indwelling catheter in the last seven days prior to the assessment reference date on the target assessment. In this case, low-risk refers to residents who do not have preexisting conditions, such as neurogenic bladder or obstructive uropathy, which predispose catheter use. This measure is based on data from the Minimum Data Set (MDS) 3.0 OBRA, PPS, and/or discharge assessments during the selected quarter. Long-stay nursing home residents are identified as those who have had 101 or more cumulative days of nursing home care.

1b.1. Developer Rationale: Significance to residents:

Catheter use is important to address because of its impact on residents. A study by Hu et al. (2014) found that inappropriate catheter use was associated with corollary outcomes of longer hospital stays, increased rate of catheterization at the time of discharge, higher incidence of UTI, and decrease in ADL function (Hu et al., 2014). Hollingsworth et al. (2013) identified additional corollary outcomes of indwelling catheterization. In a meta-analysis of 37 studies involving 2,868 adults, the authors found increased risk of urine leakage in both short- (10.6%) and long-term (52.1%) catheterization. The authors also identified a high risk of urethral strictures (3.4%) among individuals with short-term catheters. Among individuals with spinal cord injuries, gross hematuria (13.5%) and bladder cancer (1.0%) were associated with indwelling catheter use. Although this meta-analysis includes a range of patient types, not limited only to nursing home residents, the findings show that limiting catheter use may be important for preventing a host of serious health issues (Hollingsworth et al., 2013).

Gaps in performance in nursing homes:

Several studies have found that structural characteristics of nursing homes can have an impact on catheter use. Castle and Anderson (2011) found that improvements of structural characteristics, such as increasing RN staffing levels and decreasing RN turnover, significantly decreased catheter use, thus

improving resident outcomes. Nursing homes with more educated staff, including a Director of Nurses with at least a bachelor's degree, were found to have lower catheterization rates than those with less educated staff (Damkoehler, 2014). Furthermore, providers with more committed staff through a stronger culture, more experienced staff, and proper staff to resident/patient ratios have also been presented as characteristics that may be linked to adherence with infection control guidelines/practices and appropriate catheter use (Flodgren et al., 2013).

Clifton et al. (2018) studied a quality improvement initiative across Veterans Administration (VA) hospitals and found that use of best practices for catheter insertion, indication, and assessments were correlated with an improved catheter-associated urinary tract infection (CAUTI) outcome (Clifton et al., 2018). Additional processes such as a thorough and informed assessment of the resident, and evaluation of the medical need for a catheter, can sometimes decrease or prevent the use of catheters and the risks associated with their use.

Gaps in performance among specific groups of nursing home residents:

A few studies found an empirical association between social risk factors that could be measured by items available in the MDS 3.0 and catheter use, but did not offer a conceptual basis for understanding how the inherent characteristics of the social risk factor (gender, age, Medicaid coverage, and race/ethnicity) would affect catheter use.

One older study (Rogers et al., 2008) discusses a conceptual basis for using gender as a clinical risk adjuster. However, it is possible that the association between gender and catheter use is reflecting other gender-specific conditions for which facilities should be held accountable when providing care. Similarly, studies looked at the association between age and risk of catheterization and found that older adults are more likely to be catheterized due to a higher likelihood of experiencing adverse health outcomes when being transferred from the acute care setting to SNFs (Burke et al., 2018).

We examined the percentage of long-stay residents with each social risk factor identified in the literature as having an empirical association with catheter use, compared to those without that social risk factor, and used Chi-Squared tests to determine whether these differences were statistically significant.

While all differences were statistically significant and absolute percentages are small, the differences across most subpopulations are relatively large. Among residents who are eligible for Medicaid, 2.0% have/had a catheter inserted and left in their bladder and, among those ineligible for Medicaid, 3.0% have/had a catheter inserted (?2(1) = 472.3, $p < 0.001$). For residents aged 85 years or older, 1.8% have/had a catheter inserted and left in their bladder, compared to 2.4% of younger residents (?2(1) = 474.8, $p < 0.001$). Whereas 3.3% of the male residents have/had a catheter inserted and left in their bladder, 1.6% of the female residents have/had a catheter inserted (?2(1) = 3,200, $p < 0.001$). Alternatively, the difference across race/ethnicity subpopulations is relatively small. Among residents who are non-Hispanic white, 2.1% have/had a catheter inserted and left in their bladder and, among those who are non-white, 2.3% have/had a catheter inserted (?2(1) = 33.7, $p < 0.001$).

Importance to stakeholders:

On May 23, 2019, RTI International convened a web-based technical expert panel (TEP) meeting to obtain expert input on future directions for measure development and maintenance of quality measures for nursing homes based on the Minimum Data Set 3.0. In the pre-TEP survey, 6 out of 10 TEP members rated this measure as "very important" (scoring it a 4 or 5 out of a scale from 1–5), according to the following criteria : is an established priority area (National Quality Strategy); addresses a demonstrated high-impact aspect of health care (e.g., affects large numbers); has external evidence of importance,

such as consensus standards; and has evidence of disparities for the quality domain. The majority of TEP members explicitly affirmed the face validity of NQF #0686.

Most TEP members viewed the measure as important because it kept attention on removing catheters and reducing catheter use where possible, given the relationship between catheterization and adverse health outcomes and reduced quality of life. TEP members concurred that measuring catheterization is important and that the measure encourages facilities to focus on continence care. Several TEP members acknowledged that this measure supports quality improvement, with one TEP member noting that this QM encourages facilities to focus on reducing catheterization that is done for convenience or without clinical justification (RTI International, 2019).

Castle, N. G., Anderson, R. A. (2011). Caregiver staffing in nursing homes and their influence on quality of care: using dynamic panel estimation methods. Medical Care, 49(6), 545-552

Clifton, M., Kralovic, S. M., Simbartl, L. A., Minor, L., Hasselbeck, R., Martin, T., & Roselle, G. A. (2018). Achieving balance between implementing effective infection prevention and control practices and maintaining a home-like setting in US Department of Veterans Affairs nursing homes. American journal of infection control, 46(11), 1307-1310.

Damkoehler, G. Using a nurse-led protocol to reduce infections. McKnight's. Oct. 15, 2014. Available from: http://www.mcknights.com/marketplace/using-a-nurse-led-protocol-to-reduce-infections/article/377448/

Flodgren, G., Conterno, L. O., Mayhew, A., Omar, O., Pereira, C. R., & Shepperd, S. (2013). Interventions to improve professional adherence to guidelines for prevention of device-related infections. Cochrane Database of Systematic Reviews, (3).

Hollingsworth, J. M., et al. (2013). "Determining the Noninfectious Complications of Indwelling Urethral Catheters." Ann Intern Med. 159(6): 401-10

Hu, F. W., et al. (2014). "Inappropriate use of urinary catheters among hospitalized elderly patients: Clinician awareness is key." Geriatr Gerontol Int. Dec. 11.

RTI International. (2019, June). Technical Expert Panel Summary Report: Maintenance of Nursing Home Quality Measures Prepared under CMS Contract No. HHSM-500-2013-13015I.

RTI analysis of Q3, 2018 MDS 3.0 Data (programming reference: LJC67/LJC09_request_686_31_32_rev.log)

S.4. Numerator Statement: The numerator is the number of long-stay nursing home residents in the denominator sample with an episode during the selected quarter with a target assessment that indicates the use of indwelling catheters within the last seven days.

S.6. Denominator Statement: The denominator includes all long-stay residents in the nursing home who have an episode during the selected quarter with a qualifying target assessment (OBRA, PPS, or discharge assessment) and who do not meet the exclusion criteria.

S.8. Denominator Exclusions: The denominator exclusion criteria for this quality measure are as follows: 1) The target assessment is an admission assessment, a PPS 5-day assessment or a PPS readmission/return assessment; 2) The target assessment indicates that indwelling catheter status is missing; 3) The target assessment indicates neurogenic bladder or neurogenic bladder status is missing; or 4) The target assessment indicates obstructive uropathy or obstructive uropathy status is missing.

De.1. Measure Type: Outcome

S.17. Data Source: Assessment Data

S.20. Level of Analysis: Facility

IF Endorsement Maintenance – Original Endorsement Date: Mar 03, 2011 **Most Recent Endorsement Date:** Mar 03, 2011

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

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

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

Preliminary Analysis: Maintenance of Endorsement

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

Criteria 1: Importance to Measure and Report

1a. Evidence

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

1a. Evidence. The evidence requirements for a *structure, process or intermediate outcome* measure is that it is based on a systematic review (SR) and grading of the body of empirical evidence where the specific focus of the evidence matches what is being measured. For measures derived from patient report, evidence also should demonstrate that the target population values the measured process or structure and finds it meaningful.

The developer provides the following evidence for this measure:

- **Systematic Review of the evidence specific to this measure?** ☐ **Yes** ☒ **No**
- **Quality, Quantity and Consistency of evidence provided?** ☐ **Yes** ☒ **No**
- **Evidence graded?** ☐ **Yes** ☒ **No**

Summary of prior review

The developer provided a logic model linking nursing home structure to the process of placement of a urinary cathether. In addition, there is evidence that longer term catheter use is associated higher rates of catheter-associated urinary tract infections (CAUTI), an outcome that is associated with significant morbidity and mortality. The developer provided general guidelines that suggest with good evidence (category 1B - A strong recommendation supported by low quality evidence suggesting net clinical benefits or harms or an accepted practice [e.g., aseptic technique] supported by low to very low quality evidence) that urinary catheter should only be used when absolutely needed, and that they should not be routinely used in nursing homes (the setting of this measure), or during operative procedures

routinely, and that when they are needed their use should be minimized. There was no systematic review of the evidence on this measure.

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 provided some new evidence that supports the rationale of this measure that certain structural measures (particularly increasing RN staffing) is associated with decreased indwelling catheter use. In addition, there was some new evidence of "gap" particularly in hospitals where there continues to be inappropriate use of catheters in hospitals.

Questions for the Committee:

If the developer provided updated evidence for this measure:

- Does the Committee agree there is no need for repeat discussion and vote on Evidence?
- Is there a concern that there was no systematic review of the evidence for this measure?

Guidance from the Evidence Algorithm

Process measures not based on systematic review (Box 3) -> Empricial evidence submitted without systematic review and grading of the evidence (Box 7) -> Empirical evidence only includes all studies (Box 8) -> High certainty that benefits outweight undesirable effects -> Moderate.

The highest possible rating is Moderate.

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

RATIONALE:

There were guidelines provided that did support the concept of this measure rated with Class 1B evidence.

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 presented the distribution of facility-level scores on this quality measure in Quarter 3, 2018. Overall, 1,041,783 long-stay residents in 14,374 nursing homes are included in the analysis. The national facility-level mean score for this measure in Quarter 3, 2018 was 2.2% and the median score was 1.7%, suggesting a slight positive skew. The interquartile range for this measure was 3.4% and 35.2% of facilities had a perfect score of 0.0%.

Disparities

The developer did report significant disparities data for this measure. While the absolute percentages of triggering the numerator are small among each resident characteristic, the differences between most

subpopulations within a disparities category are relatively large. For residents aged 85 years or older, 1.8% have/had a catheter inserted and left in their bladder, compared to 2.4% of younger residents (?2(1) = 474.8, p < 0.001). Whereas 3.3% of the male residents have/had a catheter inserted and left in their bladder, 1.6% of the female residents have/had a catheter inserted (?2(1) = 3,200, p < 0.001).Among residents who are eligible for Medicaid, 2.0% have/had a catheter inserted and left in their bladder and, among those ineligible for Medicaid, 3.0% have/had a catheter inserted (?2(1) = 472.3, p < 0.001). Alternatively, the difference across race/ethnicity subpopulations is relatively small. Among residents who are non-Hispanic white, 2.1% have/had a catheter inserted and left in their bladder and, among residents who are non-white, 2.3% have/had a catheter inserted (?2(1) = 33.7, p < 0.001).

Questions for the Committee:

- Is there a gap in care that warrants a national performance measure?
- Are disparities considerable enough that would warrant risk adjustment or stratification?

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

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

1a. Evidence
Comments:
**no
**This measure was last endorsed in 2011. The sponsor cited a number of recent studies, suggesting that longer term catheter use is still associated with higher rates of CAUTI, an infection that can result in significant morbidity and mortality. So the updated evidence supports the need to maintain such an outcome measure for improving residents' health outcomes and quality of life.
**Evidence to support this measure is strong and directly relates to what is being measured - overuse of urinary catheters is linked to many bad outcomes, so reducing their use whenever possible will improve care. This does not measure all of the other issues that improve cath use (better nurse ratios, well educated nurses, low turnover of staff. But cath use might be a proxy for those issues.
**The evidence is primarily through guidelines and and looking at the epidemiology and correlation between staffing and catheter use. No significant new data that changes the evidence base for this submission.
**no
**The developer provided some new evidence that supports the rationale of this measure that certain structural measures (particularly increasing RN staffing) is associated with decreased indwelling catheter use. In addition, there was some new evidence of "gap" particularly in hospitals where there continues to be inappropriate use of catheters in hospitals. Since many patients are admitted to nursing homes from hospitals, this would suggest that the use of sometimes inappropriate catheters is carried over in the nursing home setting. The developer provided some evidence that structural issues (staffing) contribute to differences in catheter use. Nursing homes are required to strive to maintain residents' highest practicable level of functioning. Therefore, the inappropriate use of catheters interferes with optimal functioning and negatively affects quality of life. I am not aware of any new information that changes the evidence base. However, updates to

CMS requirements for nursing homes that have taken effect in the last two years further emphasize the goal of person-centered care in a homelike environment
**There appears to be a moderate amount of evidence to support this outcome-based quality measure which is tangentially related to the population and measure itself. There is certainly evidence that indwelling urinary catheters have been associated with adverse health outcomes and quality of life, however there does not appear to be an evidence base in this specific patient population (long term care), and there are no high quality studies/systematic reviews.

1b. Performance Gap
Comments:
**yes
** The developer presented the distribution of facility-level scores on this measure in Quarter 3, 2018. The analysis included 1,041,783 long-term residents in 14,374 nursing homes. The national mean score was 2.2% with a median of 1.7%. The interquartile range for this measure was 3.4% and 35.2% of facilities had a perfect score of 0.0%. Overall, just above one-third (38.1%) of facilities were significantly different from the national mean in Quarter 3, 2018, indicating that there are meaningful differences in facility-level scores for this measure and there is an opportunity for improvement.
** Performance variability is small but the TEP and evidence indicates that this is an important measure that warrants national performance measure.
** Yes, there does seem to variability and potential opportunity for improvement
** yes a gap-still room for improvement
** Performance data was provided; there is a performance gap between subpopulations that warrants the continuation of the measure.
**Based on the data provided, there does appear to be a persistent performance gap. The size of the population (over 1M long term care residents at over 14K facilities included in both the Q4 2017 and Q3 2018 analysis) nursing home residents also helps contextualize the small percentage points (which is a concern addressed later about overall strong performance).

Disparities:
Comments:
**yes
** To assess disparities, the developer grouped residents into different subpopulation categories, based on age, sex, Medicaid eligibility, and race/ethnicity. The developer found that the differences in the percentages of residents having a catheter inserted and left in their bladders are relatively large for subpopulations within a disparity category, indicating significant disparities for this measure.
** Yes - there were disparities in care based on age and gender, but "relatively small" differences based on race/ethnicity.
** Yes, data on disparaties was discussed. There were some differences in usage by gender and dual medicaid beneficiary status.
**yes
** While the absolute percentages of triggering the numerator are small among each resident characteristic, the differences between most 6 subpopulations within a disparities category are relatively large. For residents aged 85 years or older, 1.8% have/had a catheter inserted and left in their bladder, compared to 2.4% of younger residents (?2(1) = 474.8, p < 0.001). Whereas 3.3% of the male residents have/had a catheter inserted and left in their bladder, 1.6% of the female residents have/had a catheter inserted (?2(1) = 3,200, p < 0.001).Among residents who are eligible for Medicaid, 2.0% have/had a catheter inserted and left in their bladder and, among those ineligible for Medicaid, 3.0% have/had a catheter inserted (?2(1) = 472.3, p < 0.001). Alternatively, the difference

across race/ethnicity subpopulations is relatively small. Among residents who are non-Hispanic white, 2.1% have/had a catheter inserted and left in their bladder and, among residents who are non-white, 2.3% have/had a catheter inserted (?2(1) = 33.7, p < 0.001)
**Disparities between patients based on insurance type, age and gender appear to exist. This however is complicated by the evidence provided from the AMDA clinical practice guideline (p35) that indicates that urinary incontinence is more common is women and as patients page. This raises the question around if the gap in performance is confounded by indication.

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: Patient Safety project team staff
Evaluation of Reliability and Validity: Link A (Project Team staff)

Reliability

- Critical Data Element Reliability
    - For the bladder and bowel continence items, including catheter use, the average kappa for gold-standard nurse to gold-standard nurse agreement was 0.949, and the average kappa for gold-standard nurse to facility nurse agreement was 0.945. For the MDS items for pressure ulcers, average kappa for gold-standard to gold-standard nurse agreement was 0.905, and the average kappa for gold-standard nurse to facility nurse agreement

was 0.937. All values for kappa are well above the value of 0.60, which is generally considered substantial agreement (Saliba & Buchanan, 2008).

- Performance Measure Score Reliability
    - Signal-to-noise: The signal-to-noise ratio for this measure was 0.132 ($p < 0.001$) indicating that 13.2% of the variance in scores for this measure in Quarter 3, 2018 was explained by inter-facility characteristics (including the underlying quality of care in each facility) (RTI International, 2019a). Thus, this measure is somewhat reliable in separating facility characteristics from the noise of population variance.
    - Split-half reliability analysis: Correlations above 0.6 are generally considered as evidence of strong reliability (Armitage & Berry, 1994; Bland & Altman, 1986). The split-half correlation for this measure was positive, but the relationship was weak ($r = 0.26$, œÅ = 0.22, $p < .001$), and the ICC was 0.27 ($p < .001$), providing limited evidence of internal reliability (RTI International, 2019b). These low correlations were expected due to only a small amount of variation in performance among providers. The national-level distribution indicates one-third of all providers achieved a perfect score of 0.0% (lower scores demonstrate higher quality), giving rise to a positively-skewed distribution of provider performance; perhaps even more significant, the range of non-perfect scores was very small, between 1.7% in the 50th percentile and 5.5% in the 90th percentile (see Table 11 in Section 2b4.2 where these data are presented. Taking this finding into account, intra-facility splits of the data would result in pairs of zero values (0.0%) being compared for one-third of all facilities. Since correlations are calculated using the covariance of the data, and the individual variances to naturalize the covariance to report a value between -1 and 1, the small amount of variance in performance was expected to yield low correlation coefficients. Further, given the overall high-performance scores among non-perfect providers and the variation in denominator sizes, it's likely that splitting the data would also result in scores of 0% being compared to non-zero values. This may also result in low correlations because the data may not vary in a systematic or predictable way across the split samples.

Validity

- Performance Measure Score Validity
    - Correlation with related quality measures: Among facilities who could report both measures, RTI calculated the correlation between the facility's percentile rank on NQF #0686 (Percent of Residents Who Have/Had a Catheter Inserted and Left in Their Bladder (Long Stay)) and the facility's percentile rank on NQF #0684 (Percent of Residents with a Urinary Tract Infection (Long Stay)) and found a positive, but weak ($\rho = 0.110$), and statistically significant ($p < 0.001$) correlation. Among facilities who could report both measures, RTI also calculated the correlation between the facility's percentile rank on NQF #0686 (Percent of Residents Who Have/Had a Catheter Inserted and Left in Their Bladder (Long Stay)) and NQF #0685 (Percent of Low Risk Residents Who Lose Control of their Bowels or Bladder (Long Stay)) and found a negative, but weak ($\rho = -0.006$) and statistically insignificant relationship.
    - Variation by State: RTI conducted a one-way analysis of variance (ANOVA) and examined the interquartile range in mean state-level scores across states to assess whether state characteristics were a source of facility measure score variation for NQF #0686. The

proportion of variance in this measure explained by the state in which facilities are located is 0.32% ($p < 0.001$). The interquartile range of mean state-level scores is 3.4% (RTI International, 2019b).

- Seasonality: RTI examined the national-level mean and median quality measure scores for each quarter from Quarter 1, 2011, to Quarter 3, 2018. The results are presented in Figure 1 in the MIF below. The national-level means and medians have both decreased almost monotonically since Quarter 1 of 2011. These results show no evidence of seasonal variation. Further, the decreasing trend indicates that facilities may have improved their ability to assess residents during this period, for example, by offering individualized treatment and services to achieve or maintain normal elimination function.
- Stability analysis: Figure 2 illustrates the changes in facility rank by quality measure score from Quarter 2, 2018, to Quarter 3, 2018 ($k = 14,261$). The majority (58.14%) of facilities ranked within the same decile in both quarters. Shifts of more than 3 deciles were less common, occurring for approximately 23.73% of facilities. Thus, both facility scores and relative ranks for this measure are stable from one quarter to the next.
- Confidence interval analysis: Another measure of validity is performance relative to the mean: high-performing facilities should have scores that are significantly below-average, and low-performing facilities should be significantly above-average. Table 2 shows the proportions of facilities that scored significantly higher or lower (i.e., different) than the national facility-level mean in Quarter 3, 2018. 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.
- In general, there were fewer facilities with quality measure scores that were statistically significantly ($p \leq .05$) higher than the national mean of 2.2% (1.54%) than those with scores that were statistically significantly lower than the national mean (36.55%), indicating that more facilities perform better (lower scores are better) than the national facility-level mean.
- The proportions of facilities with scores that are significantly different from the national mean vary as a function of the number of residents included in the denominator for this measure; the percentage of facilities which have scores that are statistically significantly different from the mean generally decreases as number of residents increases.
- Overall, just above one-third (38.10%) of facilities were significantly different from the national mean in Quarter 3, 2018, suggesting that there are some meaningful differences in facility-level scores for this measure and providing evidence of validity for NQF #0686.
- Average change in performance across years: Table 3 in the MIF below presents the changes in provider performance scores from year to year, using fiscal year (FY) 2013 – 2018. On average, provider scores changed by less than 0.01 percentage points on NQF #0686 between years. Few facilities experienced a change in performance by 0.05 or greater and over 90% of provider scores changed by 0.03 points or less between years. The mean nursing home score change between FY 2017 and FY 2018 (the coding

guideline changed at the beginning of FY 2018) was 0.003, which is similar to previous mean facility-score changes between other years where there was no change in clinical coding guidelines (0.001 – 0.005). While impacts of the change that occurred on October 1st, 2015 with the implementation of the ICD-10 codes will continue to be monitored as more data become available, there is no indication that there was a significant impact on nursing home performance due to the change in clinical coding guidelines. Thus, the output suggests that changes to the clinical coding guidelines did not have a substantial effect on provider performance and do not appear to be a threat to the validity of NQF #0686.

- Face validity: The majority of TEP members explicitly affirmed the face validity of NQF #0686. The TEP supported currently specified exclusion criteria of neurogenic bladder and obstructive uropathy, as well as the risk adjustment covariates of pressure ulcers at stages II – IV and frequent bowel incontinence. Some TEP members suggested including additional risk-adjustors but were countered by other members of the TEP, indicating no strong consensus among TEP members. Other TEP members explained the trade-offs of including additional risk adjustors that may not be well documented in the MDS and how that may compromise the validity of the current measure. Overall, TEP members voiced support for the face validity of NQF #0686 as it is currently specified (RTI International, 2019).

Questions for the Committee regarding reliability:

- Do you have any concerns that the measure can be consistently implemented (i.e., are measure specifications adequate)?

Questions for the Committee regarding validity:

- Do you have any concerns regarding the validity of the measure (e.g., exclusions, risk-adjustment approach, etc.)?

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

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

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

2a1. Reliability – Specifications

Comments:

**none

** Appropriate reliability testing was conducted. But the signal to noise and the split-half reliability tests showed that only a small percent of the variance come from the differences among facilities.

** No concerns but there are only small variations among providers, with 1/3 having score of zero -- is it possible that the same findings exist for this measure an 0684? seems to be very similar information. Also, same findings re changes over time: on average, provider scores changed by less than 0.01 percentage points from 2013-2018. As with the measure 0684, would be good to get clarity on how the 101 days are counted and the impact of having a hospitalization (esp one that included

insertion of a catheter) - they say hospitalization doesn't reset to 0 days, but not sure how it affects the measure otherwise.

** I have some concerns about the robustness of risk adjustment factors. The variation of more than 3 deciles between Q2 and Q3 for 24% of facilities suggests factors other than quality may be impacting catheter usage. I would like to see the rsquare of the model with the riska djustment factors to better understand what amount of variability is explained by the two risk adjustment factors used.

** I do not have concerns

** I have no concerns about the measure being consistently implemented based on the information provided.

**The data elements (numerator, denominator and exclusions) are clearly defined. The risk adjustment criteria is clear (bowel incontinence and pressure ulcers using logistic regression). The reliability of data elements are reinforced by the strong kappas.

2a2. Reliability – Testing

Comments:

**no

**no

**no

** As above regarding confounding by unaccounted for patient factors

**no

**no

**It is challenging to assess given the small amount of variation in performance. Based on the analysis alone, it appears that the measure has moderate reliability. The signal to noise testing suggests that a small amount of the variation is accounted for by real differences in quality, and the split half reliability testing demonstrated that over 1/3 of places are already at a perfect score.

2b1. Validity – Testing

Comments:

**no

** Only about one-third (38.10%) of facilities were significantly different from the national mean in Quarter 3, 2018. This suggests that there are some meaningful differences in facility-level scores for this measure and provides evidence of validity for NQF #0686.

**no

** The corelation between the measure for catheter use and CAUTI adds credence to the validity of this metric. It would be useful to perform a chart review on a sample of LTC residents with catheters to assess how many are "preventable" or unnecessary by clinical review (not just documented indication).

**no

**no

**No- there appears to strong evidence for the validity based on the analysis provided.

2b4-7. Threats to Validity

Comments:

**none

** The developer examined the rate of missing data at both resident-level and facility-level as well as possible relationships between missing data and the scores for this measure. Overall, the mean facility-level missing data rate for this measure was quite low (0.037%), and at least 90% of facilities were not missing data on the catheter item. Rates of missing data on items used to construct this QM are very similar among individuals with selected characteristics related to catheterization (gender, age greater than or equal to 85, race/ethnicity, Medicaid eligibility, bowel incontinence, and presence of pressure ulcers at Stage II, III, or IV). Missing data is only weakly correlated with scores for this QM. Overall, missing data do not present a threat to this measure's validity. Potential impact on the validity due to changes to the clinical coding guidelines was also evaluated. The test suggests that changes to the clinical coding guidelines did not have a substantial effect on provider performance and do not appear to be a threat to the validity of this measure.

** There are meaningful differences between the highest/lowest scores. No other concerns

** It does not appear that missing data is a threat to validity. While intuitively appealing, I am not sure the data conclusively support that differences in this metric can be entirely or mostly attributed to differences in quality of care, given some of the reasons discussed earlier.

**no

**no

**One threat to validity is the overall strong performance across the country – with 1/3 of places scoring perfect and the remainder clustered in a narrow margin of strong performance.

2b2-3. Other Threats to Validity

2b2. Exclusions

2b3. Risk Adjustment

Comments:

** used statistical model

** Both exclusion and risk adjustment to bowel incontinence and pressure ulcers seem reasonable.

** The measure risk adjusts by excluding neurogenic bladder & obstructive neuropathy, and risk adjusts for pressure ulcers at stages II – IV and frequent bowel incontinence. The TEP discussed adding other adjustments but did not come to consensus

** I do not think the risk adjustment has been fully developed. It seems to be based on expert opinion rather than empirical knowledge. Burn patients and other surgical patients may need to be included in the exclusion list, although they likley make up a minority of cases. Understanding through data what conditions are assoctaied with catheter use (there was some broad data provided) would help develop more robust exclusion and risk-adjustment criteria

** risk adjusted but additional adjustments may be appropriate

** Exclusions are appropriate as described. Risk adjustment has been addressed at the start of care

**Exclusions seem appropriate, there are no social risk factors in the risk adjustment model, and the risk adjustment variables (bowel incontinence and pressure ulcers), testing and strategy seem reasonable and straightforward.

Scientific Acceptability: Preliminary Analysis Form

Measure Number: 0686

Measure Title: Percent of Residents Who Have/Had a Catheter Inserted and Left in Their Bladder (Long Stay)

Type of Measure:

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

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

Data Source:

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

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

☐ **Enrollment Data** ☐ **Other**

Level of Analysis:

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

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

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

Measure is:

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

RELIABILITY: SPECIFICATIONS

1. **Are submitted specifications precise, unambiguous, and complete so that they can be consistently implemented?** ☒ **Yes** ☐ **No**
2. **Briefly summarize any concerns about the measure specifications.**

   The data elements for the numerator and demonimator are clear and unambiguous, and there are clear criteria for exclusions. There are no concerns. There is no mention of changes since the last review.

RELIABILITY: TESTING

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

3. **Reliability testing level** ☒ **Measure score** ☒ **Data element** ☐ **Neither**

4. **Reliability testing was conducted with the data source and level of analysis indicated for this measure** ☒ **Yes** ☐ **No**

5. If score-level and/or data element reliability testing was NOT conducted or if the methods used were NOT appropriate, was **empirical <u>VALIDITY</u> testing** of <u>patient-level data</u> conducted?

   ☐ **Yes** ☐ **No**

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

For data element reliability, a national test of MDS 3.0 items examined the agreement between assessors (reliability). Gold-standard (research) nurses were recruited from community nursing facilities to participate in a national evaluation (Saliba & Buchanan, 2008) who were trained in the MDS 3.0 instrument, who then trained a facility nurse from each participating nursing facility in their home states. Residents participating in the test were selected to capture a representative sample of short- and long-stay residents. In this national test, the agreement between gold-standard nurses and between gold-standard and facility nurses on bladder and bowel continence, toileting, and catheter items was examined. There was also additional analysis on the risk adjustors, where agreement between gold-standard nurses and between gold-standard and facility nurses on items for pressure ulcers were examined. Cohen's kappas were calculated to assess item reliability. This was an appropriate approach for data element reliability.

Signal-to-noise testing and split-half reliability testing was conducted for reliability of the performance measure score, both of which both are acceptable approaches. Since the last review (in 2011), additional reliability testing at the measure score level was conducted using newer data (2018).

Submission document: Testing attachment, section 2a2.2

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

Submission document: Testing attachment, section 2a2.3

Data element reliability: For the bladder and bowel continence items, including catheter use, the average kappa for gold-standard nurse to gold-standard nurse agreement was 0.949, and the average kappa for gold-standard nurse to facility nurse agreement was 0.945. For the MDS items for pressure ulcers, average kappa for gold-standard to gold-standard nurse agreement was 0.905, and the average kappa for gold-standard nurse to facility nurse agreement was 0.937. These are are well above the value of 0.60, which is generally considered substantial agreement.

Signal-to-noise: The signal-to-noise ratio for this measure was 0.132 ($p < 0.001$). This indicates that 13.2% of the variance in scores for this measure in Quarter 3, 2018 was explained by inter-facility characteristics (including the underlying quality of care in each facility) (RTI International, 2019a). Therefore, this measure was deemed "somewhat reliable" in separating facility characteristics from the noise of population variance.

Split-half reliability analysis: The split-half correlation for this measure was positive, however the relationship was weak ($r = 0.26$, $\rho = 0.22$, $p < .001$), and the ICC was 0.27 ($p < .001$). This provided limited evidence of internal reliability. However, the developer thought that these low correlations were expected due to only a small amount of variation in performance among providers. One-third of all providers achieved a perfect score of 0.0% (lower scores demonstrate higher quality), giving rise to a positively-skewed distribution of provider performance. In addition, the range of non-perfect scores was very small, between 1.7% in the 50th percentile and 5.5% in the 90th percentile. The way the analysis was constructed (i.e. split-half reliability analysis), intra-facility splits of the data would result in pairs of zero values (0.0%) being compared for one-third of all facilities. This may have caused the low correlations.

Regardless of the developers explanations, there are some concerns over the performance score reliability, given the combined results of the signal-to-noise as well as the split-half reliability testing. This should be discussed by the committee.

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

Submission document: Testing attachment, section 2a2.2

☒ **Yes**

☐ **No**

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

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

   Submission document: Testing attachment, section 2a2.2

   ☒ **Yes**

   ☐ **No**

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

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

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

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

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

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

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

    Appropriate reliability testing was conducted; however, the signal-to-noise and the split-half reliability testing demonstrate that only a small percent of the variation are actually interfacility differences and the split-half reliability demonstrates that about 1/3 are perfect, which lead to a low correlation coefficient.

VALIDITY: ASSESSMENT OF THREATS TO VALIDITY

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

    Submission document: Testing attachment, section 2b2.

    The exclusions are clinically appropriate. Exclusions resulted in 16.3% of the potential cases being excluded. There is no mention of changes or feedback since the last review.

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

    Submission document: Testing attachment, section 2b4.

    There are many zeros for this measure with respect to performance and not a lot of variation among the remaining sites that were tested suggesting that most sites are performing well on this measure.

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

    Submission document: Testing attachment, section 2b5.

    No specific concerns. There is one data source. No changes since last review.

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

Submission document: Testing attachment, section 2b6.

No specific concerns. No changes since last review.

16. **Risk Adjustment**

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

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

☒ Yes ☐ No ☐ Not applicable

16c. **Social risk adjustment:**

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

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

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

16d.**Risk adjustment summary:**

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

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

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

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

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

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

The risk adjustment approach includes two factors: bowel incontinence and pressure ulcers. The developer has provided a reasonable justification for this. The developer does not adjust for social risk factors, and while studies have demonstrated relationships between social risk factors and this outcome, since no conceptual relationship is plausible (aside from relating to quality differences seen in different settings that may see different populations), this was not included.

For cost/resource use measures ONLY:

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

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

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

N/A

VALIDITY: TESTING

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

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

☒ **Face validity**

☒ **Empirical validity testing of the measure score**

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

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

Submission document: Testing attachment, section 2b2.2

There were several methods for assessing validity and potential threats to validity: 1) Correlation with related quality measures, 2) Variation by state, 3) Seasonality, 4) Stability analysis, 5) Confidence interval analysis, 6) Average change in performance across years and 7) Face validity. All of these approaches are thoughtful and appropriate. There were no major changes mentioned since the last review.

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

Submission document: Testing attachment, section 2b2.3

Overall, based on the results, it demonstrates validity for this measure based on the results provided by the developer.

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

Submission document: Testing attachment, section 2b1.

☒ **Yes**

☐ **No**

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

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

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

☐ **Yes**

☐ **No**

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

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

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

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

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

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

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

    The score level testing was comprehensive and robust and demonstrated convergent and face validity.

FOR COMPOSITE MEASURES ONLY: Empirical analyses to support composite construction

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

    ☐ **High**

    ☐ **Moderate**

    ☐ **Low**

    ☐ **Insufficient**

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

    N/A

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.**

    No

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.

These data are regularly collected in electronic format as part of the MDS 3.0 in all Medicare approved nursing homes.

Questions for the Committee:

- Are there any concerns about feasibility by the committee?

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

Committee Pre-evaluation Comments:
Criteria 3: Feasibility

3. Feasibility

Comments:

**none

** All data elements are in defined fields in electronic clinical data, and the general data collection method for the MDS 3.0 is currently operational and mandatory for all Medicare/Medicaid certified nursing homes; no issues are anticipated. And CMS provides coding directions for bowel and bladder items in the MDS 3.0 via the RAI Manual and other mediums. So the preliminary rating for feasibility is high.

** Seems quite feasible with little burden on the provider since it uses info from regular assessments done for all nursing home patients. These minimum data sets presumably include the presence of a catheter

** Not my area of expertise, but it appears that data collection for this metric is fairly easily available through the EMR.

**no

** Data is from the MDS data set

**No concerns

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

This quality measure 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 may access nursing home performance scores on this quality measure via the Nursing Home Compare website.

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

To ensure the continued value and efficacy of the measure, the developer convened a Technical Expert Panel (TEP) to obtain input from providers, residents, and caregivers on the importance, validity, and use of two nursing home quality measures: (1) Percentage of Residents with a Urinary Tract Infection (Long Stay) (NQF #0684); and (2) this measure (NQF #0686).

TEP members were asked to rate the importance of NQF #0686 on a scale from 1-5 (higher scores are better) based on the following criteria : is an established priority area (National Quality Strategy); addresses a demonstrated high-impact aspect of health care (e.g., affects large numbers); has external evidence of importance, such as consensus standards; and has evidence of disparities for the quality domain. 6 out of 10 TEP members rated this measure as "very important" (rating it a 4 or 5), noting that this measure is important to facilitate quality improvements in nursing homes by bringing continued attention to catheterization rates and working towards timely removal of catheters and reducing catheter use where possible. This measure was also described as tracking a critical health outcome and noted the importance of having this measure for educational opportunities so that nursing homes can take actionable improvements in the care offered to residents.

Other TEP members also noted that they use this QM to track facility performance and this measure serves as an opportunity to educate clinical staff about the impact catheter use has on residents' health outcomes and quality of life. This measure also encourages providers to focus on continence care and reduce catheterization for convenience or without clinical justification. TEP members expressed concerns that removing the measure from Nursing Home Compare would lead to an increase in unnecessary catheterizations as nursing homes would be more likely to focus their attention on other care domains that are publically reported.

TEP members also affirmed clear support for this measure's validity and there was broad support for maintaining the current risk adjustment model to avoid making the risk adjustment model too complex. The TEP members also reached consensus that there was no evidence for including any of the social risk factors in the risk adjustment model and the current exclusions are appropriate. In addition, other TEP

members explained that additional risk-adjustment could obscure areas for performance improvement in this quality domain.

Other TEP members inquired about the rates of missing data to determine if that may reduce the validity of the measure. However, missing data on any of the items used to calculate this measure are rare and the average facility missing rate of 0.037% demonstrates that missing data do not present a threat to the measure's validity. Last, one TEP member expressed concerns that the measure "…could be construed as somewhat burdensome" because facilities might not modify an MDS assessment for a long-stay resident when a catheter is removed. Although data element validity warrants continued monitoring and this may be an opportunity for provider training, there was no evidence of this found during the MDS 3.0 pilot testing of these items. Other TEP members also pointed out that this should not be an issue given the way the measure is constructed. For example, when a long-stay resident's catheter is removed, if the episode is ongoing, this will be captured on the next quarterly assessment and the resident will no longer be triggering the numerator criteria.

***Questions for the Committee**:*

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

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

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

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

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

Improvement results.

Figure A1 in the appendix of the MIF presents trend of the national mean and median for this measure across all available quarters (Q1 2011 – Q3 2018). The national facility-level mean and median scores have trended steadily downward since Q1 2011, indicating a general improvement in performance over time. The mean score for this measure was 4.39% in Quarter 1 of 2011 and the median score was 3.74%, demonstrating a positively skewed distribution from many high performing facilities. In Quarter 3 of 2018, the mean and median were 2.24% and 1.66%, respectively.

United States Medicare/Medicaid certified nursing homes with eligible long-stay residents. In Quarter 3, 2018 there were 15,241 eligible facilities containing 1,118,025 residents eligible for inclusion in the measure (before applying sample size restrictions and without excluding residents that do not have a prior and target assessment); 14,374 facilities (94.3%) containing 1,041,783 residents (93.2%) had sufficient sample sizes (20 or more long-stay residents included in the denominator) to report on this measure after applying all measure exclusion criteria and facility sample size restrictions.

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 did not report any unexpected findings.

Potential harms

The developer did not report any potential harms.

Additional Feedback:

The developer did not report any additional feedback.

***Questions for the Committee**:*

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

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

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

4a1. Use - Accountability and Transparency

Comments:

** Issue of use in long term care for paraplegic has been raised in past conversations

** The measure is currently public reported and used in the accountability program. According to the developer, 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 may access nursing home performance scores on this quality measure via the Nursing Home Compare website. The developer sought feedbacks on the measure by convening a Technical Expert Panel (TEP) to obtain input from providers, residents, and caregivers. 6 out of 10 TEP members rated this measure as "very important" (rating it a 4 or 5), noting that this measure is important to facilitate quality improvements in nursing homes by bringing continued attention to catheterization rates and working towards timely removal of catheters and reducing catheter use where possible. This measure was also described as tracking a critical health outcome and noted the importance of having this measure for educational opportunities so that nursing homes can take actionable improvements in the care offered to residents.

** The measure is used on Nursing Home Compare for long-stay residents. Fairly easy for the public to find & I think understandable to the public. This also provides feedback to the facilities as they can see how they compare to each other. The developer also convened a TEP to provide feedback on the measure. It appears to be used to improve practices. The scores have been trending down since 2011, indicating improvements. Developers said they got feedback from "residents" as part of the TEP, but the list were all providers. Would be good to see assessments from patients/consumers/public to see

how they interpret & would possibly use this measure. It might require some education of the evidence/why this measure is important, but would be worthwhile to assess what it means to the people using these facilities.

** It appears that these data are publicly reported. The feedback mechanisms aren't completely clear to me.

** I am not sure

** This measure is used as part of Nursing Home Compare.

**Public reporting likely drives performance. 10 person TEP convened with generally supportive comments and no substantive changes. It is interesting that only 6/10 rated it as a 4 or 5 on a 5 point likert scale. However qualitative comments seem to support it's use. It is not clear to me if other systematic feedback has been collected.

4b1. Usability – Improvement

Comments:

** none this measure is for low risk patients

** No harm has been identified. More importantly, benefits of this outcome measure in promoting high-quality and safe care for long-term residents outweigh any unintended negative consequences, if they exist.

** Since the overuse of catheters is found to lead to so many other health problems, this measure is definitely a benefit and a good measure of quality. Especially since the patients involved are low-risk patients.

** These data are easily translatable into performance improvement initiatives. Fortunately there should be minimal unintended consequences for this metric, except perhaps decubitus ulcers or graft failures.

** not sure developer actually measured unintended consequences but state benefits outweigh risk

** Improvement in the appropriate use of catheters would improve quality of life, decrease risk of infection, support skin health and reduce the use of supplies which are costly to facilities

**No concerns.

Criterion 5: Related and Competing Measures

Related or competing measures

The developer did not report any related or competing measures. However, there are several outcome measures related to this measure:

0138 : National Healthcare Safety Network (NHSN) Catheter-associated Urinary Tract Infection (CAUTI) Outcome Measure, Centers for Disease Control and Prevention

0684: Percent of Residents with a Urinary Tract Infection (Long-Stay), Centers for Medicare & Medicaid Services

Harmonization

The developer did not describe any need for harmonization as they did not report related measures. The committee may discuss the need for harmonization with these related measures.

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

5. Related and Competing

Comments:

** CAUTI?

** Not aware of.

** No competing measures.

** unknown. There is the catheter utilization metric through NHSN at a unit, not individual level, but I am not sure if that is a publicly reported metric for LTC

**no

**no

**Ultimately, although this is an outcome-based measure (indwelling urinary catheters) the health outcome with the strongest evidence based tied to this is UTIs. Interestingly, this is one domain for which there already appear to be measures for LTC (0684) and more specifically, tied to indwelling catheters (0138). This does beg the question if harmonization is called for.

Public and Member Comments

Comments and Member Support/Non-Support Submitted as of: 1/21/2020

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

Developer Submission

Brief Measure Information

NQF #: 0686

Corresponding Measures:

De.2. Measure Title: Percent of Residents Who Have/Had a Catheter Inserted and Left in Their Bladder (Long Stay)

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

De.3. Brief Description of Measure: This measure reports the percentage of low risk, long-stay residents who have had an indwelling catheter in the last seven days prior to the assessment reference date on the target assessment. In this case, low-risk refers to residents who do not have preexisting conditions, such as neurogenic bladder or obstructive uropathy, which predispose catheter use. This measure is based on data from the Minimum Data Set (MDS) 3.0 OBRA, PPS, and/or discharge assessments during the selected quarter. Long-stay nursing home residents are identified as those who have had 101 or more cumulative days of nursing home care.

1b.1. Developer Rationale: Significance to residents:

Catheter use is important to address because of its impact on residents. A study by Hu et al. (2014) found that inappropriate catheter use was associated with corollary outcomes of longer hospital stays, increased rate of catheterization at the time of discharge, higher incidence of UTI, and decrease in ADL function (Hu et al., 2014). Hollingsworth et al. (2013) identified additional corollary outcomes of indwelling catheterization. In a meta-analysis of 37 studies involving 2,868 adults, the authors found increased risk of urine leakage in both short- (10.6%) and long-term (52.1%) catheterization. The authors also identified a high risk of urethral strictures (3.4%) among individuals with short-term catheters. Among individuals with spinal cord injuries, gross hematuria (13.5%) and bladder cancer (1.0%) were associated with indwelling catheter use. Although this meta-analysis includes a range of patient types, not limited only to nursing home residents, the findings show that limiting catheter use may be important for preventing a host of serious health issues (Hollingsworth et al., 2013).

Gaps in performance in nursing homes:

Several studies have found that structural characteristics of nursing homes can have an impact on catheter use. Castle and Anderson (2011) found that improvements of structural characteristics, such as increasing RN staffing levels and decreasing RN turnover, significantly decreased catheter use, thus improving resident outcomes. Nursing homes with more educated staff, including a Director of Nurses with at least a bachelor's degree, were found to have lower catheterization rates than those with less educated staff (Damkoehler, 2014). Furthermore, providers with more committed staff through a stronger culture, more experienced staff, and proper staff to resident/patient ratios have also been presented as characteristics that may be linked to adherence with infection control guidelines/practices and appropriate catheter use (Flodgren et al., 2013).

Clifton et al. (2018) studied a quality improvement initiative across Veterans Administration (VA) hospitals and found that use of best practices for catheter insertion, indication, and assessments were correlated with an improved catheter-associated urinary tract infection (CAUTI) outcome (Clifton et al., 2018). Additional processes such as a thorough and informed assessment of the resident, and

evaluation of the medical need for a catheter, can sometimes decrease or prevent the use of catheters and the risks associated with their use.

Gaps in performance among specific groups of nursing home residents:

A few studies found an empirical association between social risk factors that could be measured by items available in the MDS 3.0 and catheter use, but did not offer a conceptual basis for understanding how the inherent characteristics of the social risk factor (gender, age, Medicaid coverage, and race/ethnicity) would affect catheter use.

One older study (Rogers et al., 2008) discusses a conceptual basis for using gender as a clinical risk adjuster. However, it is possible that the association between gender and catheter use is reflecting other gender-specific conditions for which facilities should be held accountable when providing care. Similarly, studies looked at the association between age and risk of catheterization and found that older adults are more likely to be catheterized due to a higher likelihood of experiencing adverse health outcomes when being transferred from the acute care setting to SNFs (Burke et al., 2018).

We examined the percentage of long-stay residents with each social risk factor identified in the literature as having an empirical association with catheter use, compared to those without that social risk factor, and used Chi-Squared tests to determine whether these differences were statistically significant.

While all differences were statistically significant and absolute percentages are small, the differences across most subpopulations are relatively large. Among residents who are eligible for Medicaid, 2.0% have/had a catheter inserted and left in their bladder and, among those ineligible for Medicaid, 3.0% have/had a catheter inserted (?2(1) = 472.3, $p < 0.001$). For residents aged 85 years or older, 1.8% have/had a catheter inserted and left in their bladder, compared to 2.4% of younger residents (?2(1) = 474.8, $p < 0.001$). Whereas 3.3% of the male residents have/had a catheter inserted and left in their bladder, 1.6% of the female residents have/had a catheter inserted (?2(1) = 3,200, $p < 0.001$). Alternatively, the difference across race/ethnicity subpopulations is relatively small. Among residents who are non-Hispanic white, 2.1% have/had a catheter inserted and left in their bladder and, among those who are non-white, 2.3% have/had a catheter inserted (?2(1) = 33.7, $p < 0.001$).

Importance to stakeholders:

On May 23, 2019, RTI International convened a web-based technical expert panel (TEP) meeting to obtain expert input on future directions for measure development and maintenance of quality measures for nursing homes based on the Minimum Data Set 3.0. In the pre-TEP survey, 6 out of 10 TEP members rated this measure as “very important” (scoring it a 4 or 5 out of a scale from 1–5), according to the following criteria : is an established priority area (National Quality Strategy); addresses a demonstrated high-impact aspect of health care (e.g., affects large numbers); has external evidence of importance, such as consensus standards; and has evidence of disparities for the quality domain. The majority of TEP members explicitly affirmed the face validity of NQF #0686.

Most TEP members viewed the measure as important because it kept attention on removing catheters and reducing catheter use where possible, given the relationship between catheterization and adverse health outcomes and reduced quality of life. TEP members concurred that measuring catheterization is important and that the measure encourages facilities to focus on continence care. Several TEP members acknowledged that this measure supports quality improvement, with one TEP member noting that this QM encourages facilities to focus on reducing catheterization that is done for convenience or without clinical justification (RTI International, 2019).

Castle, N. G., Anderson, R. A. (2011). Caregiver staffing in nursing homes and their influence on quality of care: using dynamic panel estimation methods. Medical Care, 49(6), 545-552

Clifton, M., Kralovic, S. M., Simbartl, L. A., Minor, L., Hasselbeck, R., Martin, T., & Roselle, G. A. (2018). Achieving balance between implementing effective infection prevention and control practices and maintaining a home-like setting in US Department of Veterans Affairs nursing homes. American journal of infection control, 46(11), 1307-1310.

Damkoehler, G. Using a nurse-led protocol to reduce infections. McKnight's. Oct. 15, 2014. Available from: http://www.mcknights.com/marketplace/using-a-nurse-led-protocol-to-reduce-infections/article/377448/

Flodgren, G., Conterno, L. O., Mayhew, A., Omar, O., Pereira, C. R., & Shepperd, S. (2013). Interventions to improve professional adherence to guidelines for prevention of device-related infections. Cochrane Database of Systematic Reviews, (3).

Hollingsworth, J. M., et al. (2013). "Determining the Noninfectious Complications of Indwelling Urethral Catheters." Ann Intern Med. 159(6): 401-10

Hu, F. W., et al. (2014). "Inappropriate use of urinary catheters among hospitalized elderly patients: Clinician awareness is key." Geriatr Gerontol Int. Dec. 11.

RTI International. (2019, June). Technical Expert Panel Summary Report: Maintenance of Nursing Home Quality Measures Prepared under CMS Contract No. HHSM-500-2013-13015I.

RTI analysis of Q3, 2018 MDS 3.0 Data (programming reference: LJC67/LJC09_request_686_31_32_rev.log)

S.4. Numerator Statement: The numerator is the number of long-stay nursing home residents in the denominator sample with an episode during the selected quarter with a target assessment that indicates the use of indwelling catheters within the last seven days.

S.6. Denominator Statement: The denominator includes all long-stay residents in the nursing home who have an episode during the selected quarter with a qualifying target assessment (OBRA, PPS, or discharge assessment) and who do not meet the exclusion criteria.

S.8. Denominator Exclusions: The denominator exclusion criteria for this quality measure are as follows: 1) The target assessment is an admission assessment, a PPS 5-day assessment or a PPS readmission/return assessment; 2) The target assessment indicates that indwelling catheter status is missing; 3) The target assessment indicates neurogenic bladder or neurogenic bladder status is missing; or 4) The target assessment indicates obstructive uropathy or obstructive uropathy status is missing.

De.1. Measure Type: Outcome

S.17. Data Source: Assessment Data

S.20. Level of Analysis: Facility

IF Endorsement Maintenance – Original Endorsement Date: Mar 03, 2011 **Most Recent Endorsement Date:** Mar 03, 2011

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

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

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

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

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

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

NQF_0686_Catheter_Evidence_Form_Final_10-31-19_508-637081412439300840.docx

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

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

1a. Evidence (subcriterion 1a)

NATIONAL QUALITY FORUM—Evidence (subcriterion 1a)

Measure Number (*if previously endorsed*)**:** 0686

Measure Title: Percent of Residents Who Have/Had a Catheter Inserted and Left in Their Bladder (long stay)

IF the measure is a component in a composite performance measure, provide the title of the Composite Measure here: Click here to enter composite measure #/ title

Date of Submission: 10/31/2019

Instructions

- ***Complete 1a.1 and 1a.2 for all measures. If instrument-based measure, complete 1a.3.***
- ***Complete EITHER 1a.2, 1a.3 or 1a.4 as applicable for the type of measure and evidence.***
- ***For composite performance measures:***
  - ***A separate evidence form is required for each component measure unless several components were studied together.***
  - ***If a component measure is submitted as an individual performance measure, attach the evidence form to the individual measure submission.***
- **All information needed to demonstrate meeting the evidence subcriterion (1a) must be in this form. An appendix of *supplemental* materials may be submitted, but there is no guarantee it will be reviewed.**
- **If you are unable to check a box, please highlight or shade the box for your response.**
- **Contact NQF staff regarding questions. Check for resources at Submitting Standards webpage.**

Note: The information provided in this form is intended to aid the Standing Committee and other stakeholders in understanding to what degree the evidence for this measure meets NQF's evaluation criteria.

1a. Evidence to Support the Measure Focus

The measure focus is evidence-based, demonstrated as follows:

- **Outcome: [3] Empirical data demonstrate a relationship between the outcome and at least one healthcare structure, process, intervention, or service. If not available, wide variation in performance can be used as evidence, assuming the data are from a robust number of providers and results are not subject to systematic bias.**
- **Intermediate clinical outcome: a systematic assessment and grading of the quantity, quality, and consistency of the body of evidence [4] that the measured intermediate clinical outcome leads to a desired health outcome.**
- **Process: [5] a systematic assessment and grading of the quantity, quality, and consistency of the body of evidence [4] that the measured process leads to a desired health outcome.**
- **Structure: a systematic assessment and grading of the quantity, quality, and consistency of the body of evidence [4] that the measured structure leads to a desired health outcome.**
- **Efficiency: [6] evidence not required for the resource use component.**
- **For measures derived from patient reports, evidence should demonstrate that the target population values the measured outcome, process, or structure and finds it meaningful.**
- **Process measures incorporating Appropriate Use Criteria: See NQF's guidance for evidence for measures, in general; guidance for measures specifically based on clinical practice guidelines apply as well.**

Notes

3. Generally, rare event outcomes do not provide adequate information for improvement or discrimination; however, serious reportable events that are compared to zero are appropriate outcomes for public reporting and quality improvement.

4. The preferred systems for grading the evidence are the Grading of Recommendations, Assessment, Development and Evaluation (GRADE) guidelines and/or modified GRADE.

5. Clinical care processes typically include multiple steps: assess → identify problem/potential problem → choose/plan intervention (with patient input) → provide intervention → evaluate impact on health status. If the measure focus is one step in such a multistep process, the step with the strongest evidence for the link to the desired outcome should be selected as the focus of measurement. Note: A measure focused only on collecting PROM data is not a PRO-PM.

6. Measures of efficiency combine the concepts of resource use and quality (see NQF's Measurement Framework: Evaluating Efficiency Across Episodes of Care; AQA Principles of Efficiency Measures).

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

Outcome

☒ Outcome: Catheter inserted and left in bladder

☒ Patient-reported outcome (PRO): Click here to name the PRO

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

☒ Intermediate clinical outcome (*e.g., lab value*): Click here to name the intermediate outcome

☐ Process: Click here to name what is being measured

☐ Appropriate use measure: Click here to name what is being measured

☐ Structure: Click here to name the structure

☐ Composite: Click here to name what is being measured

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

Outcomes and corollary outcomes

This outcome-based quality measure reports the percentage of long-stay nursing home residents who have had a catheter inserted and left in their bladder. Catheter use is important to address because of its impact on functioning and activities of daily living (ADL) as well as overall quality of life, and other adverse health outcomes associated with use of catheter. Catheter use could lead to other serious corollary health outcomes, such as a higher incidence of urinary tract infections (UTI), and genitourinary complications, including risk of urethral strictures and risk of urine leakage (Hollingsworth et al., 2013).

A study by Hu et al. (2014) explored the use of urinary catheters among hospitalized older adults (age 65 and older), finding that more than half (52.8%) of the urinary-catheter days observed in the study were "inappropriate" for hospitalized seniors. The most common population to experience unnecessary catheterization was women, and the primary reason cited was "convenience of care." Inappropriate catheter use was associated with corollary outcomes of longer hospital stays, increased rate of catheterization at the time of discharge, higher incidence of UTI, and decrease in ADL function. Furthermore, this study has important implications for nursing homes, since the increased rate of catherization at hospital discharge translates to more catheterized nursing home residents, which increases their risk for infection (Hu et al., 2014).

Decreasing the rate of indwelling catheterization so it is used only when appropriate is important because adverse corollary outcomes go beyond the increased incidence of infection. Hollingsworth et al. (2013) identified additional corollary outcomes of indwelling catheterization. In a meta-analysis of 37 studies involving 2,868 adults, the authors found increased risk of urine leakage in both short- (10.6%) and long-term (52.1%) catheterization. The authors also identified a high risk of urethral strictures (3.4%) among individuals with short-term catheters. Among individuals with spinal cord injuries, gross hematuria (13.5%) and bladder cancer (1.0%) were associated with indwelling catheter use. The authors conclude that these findings demonstrate that these other conditions are as prevalent as UTI. Although this meta-analysis includes a range of patient types, not limited only to nursing home residents, the findings show that limiting catheter use may be important for preventing a host of serious health issues (Hollingsworth et al., 2013).

The structural characteristics of nursing homes and the process of care delivered in nursing homes have been shown to be related to the use of catheter among their residents, or more generally associated with quality of care that impacts health outcomes of residents (Castle & Anderson, 2011; Hyer et al., 2011; Mukamel et al., 2008). ***Figure 1*** below illustrates the key structures, processes, and outcomes that are associated with lower rates of catheterization. The structures and processes listed in the figure are not exhaustive but are intended as examples.

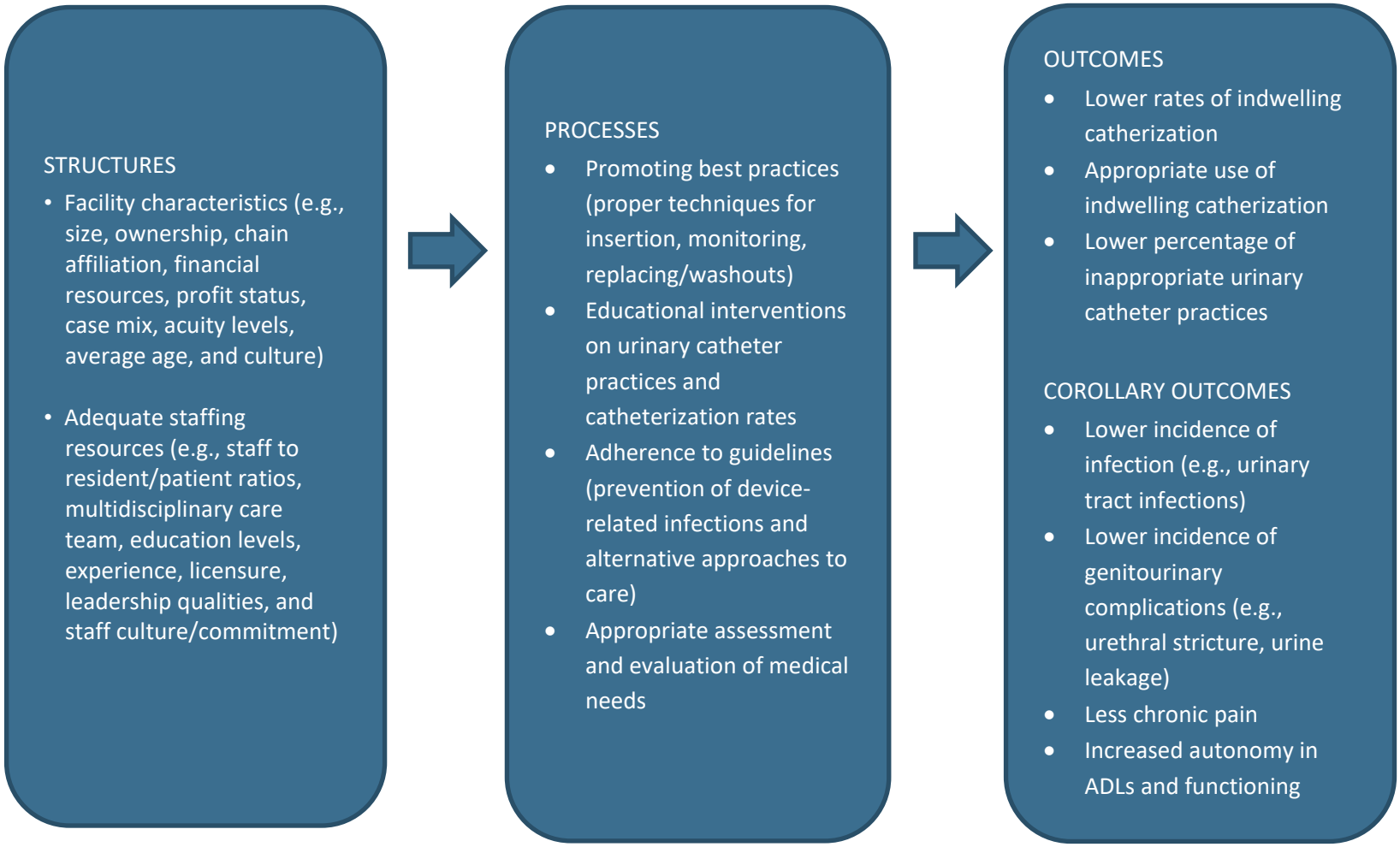
Evidence for link between structure and quality of care outcomes

Castle and Anderson (2011) found that improvements in structural characteristics such as increasing RN staffing levels and decreasing RN turnover significantly decreased catheter use, thus improving resident outcomes. Nursing homes with more educated staff, including a Director of Nurses with at least a bachelor's degree, were found to have lower catheterization rates than those with less educated staff (Damkoehler, 2014). Further, providers with more committed staff through a stronger culture, more experienced staff, and proper staff to resident/patient ratios have also been presented as characteristics that may be linked to adherence with infection control guidelines/practices and appropriate catheter use (Flodgren et al., 2013).

Evidence for link between processes and quality of care outcomes

One reason structural improvements, such are those referenced above, are thought to improve outcomes is because staff are more likely to be aware of and utilize best practices. Clifton et al. (2018) studied a quality improvement initiative across VA hospitals and found that use of best practices for catheter insertion, indication, and assessments were correlated with an improved catheter-associated urinary tract infection (CAUTI) outcome (Clifton et al., 2018). Additional processes such as a thorough and informed assessment of the resident, and evaluation of the medical need for a catheter, can sometimes decrease or prevent the use of catheters and the risks associated with their use.

Figure 1. Role of Nursing Home Structure and Processes in Rate of Resident Catheterization


STRUCTURES
• Facility characteristics (e.g., size, ownership, chain affiliation, financial resources, profit status, case mix, acuity levels, average age, and culture)
• Adequate staffing resources (e.g., staff to resident/patient ratios, multidisciplinary care team, education levels, experience, licensure, leadership qualities, and staff culture/commitment)
PROCESSES
• Promoting best practices (proper techniques for insertion, monitoring, replacing/washouts)
• Educational interventions on urinary catheter practices and catheterization rates
• Adherence to guidelines (prevention of device-related infections and alternative approaches to care)
• Appropriate assessment and evaluation of medical needs
OUTCOMES
• Lower rates of indwelling catherization
• Appropriate use of indwelling catherization
• Lower percentage of inappropriate urinary catheter practices
COROLLARY OUTCOMES
• Lower incidence of infection (e.g., urinary tract infections)
• Lower incidence of genitourinary complications (e.g., urethral stricture, urine leakage)
• Less chronic pain
• Increased autonomy in ADLs and functioning

Castle, N. G., Anderson, R. A. (2011). Caregiver staffing in nursing homes and their influence on quality of care: using dynamic panel estimation methods. *Medical Care, 49*(6), 545-552

Damkoehler, G. Using a nurse-led protocol to reduce infections. McKnight's. Oct. 15, 2014. Available from: http://www.mcknights.com/marketplace/using-a-nurse-led-protocol-to-reduce-infections/article/377448/

Flodgren, G., Conterno, L. O., Mayhew, A., Omar, O., Pereira, C. R., & Shepperd, S. (2013). Interventions to improve professional adherence to guidelines for prevention of device-related infections. Cochrane Database of Systematic Reviews, (3).

Hyer, K., Thomas, K. S., Branch, L. G., Harman, J. S., Johnson, C. E., & Weech-Maldonado, R. (2011). The influence of nurse staffing levels on quality of care in nursing homes. *The Gerontologist, 51*(5), 610-616.

Mukamel, D. B., Weimer, D. L, Spector, W. D., Ladd, H., & Zinn, J. S. (2008). Publication of quality report cards and trends in reported quality measures in nursing homes. *Health Services Research, 43*(4), 1244-1262.

Rogers, M. A., Mody, L., Kaufman, S. R., Fries, B. E., McMahon, L. F., & Saint, S. (2008). Use of urinary collection devices in skilled nursing facilities in five states. *Journal of the American Geriatrics Society*, *56*(5), 854-861.

Hu, F. W., et al. (2014). "Inappropriate use of urinary catheters among hospitalized elderly patients: Clinician awareness is key." Geriatr Gerontol Int. Dec. 11.

Hollingsworth, J. M., et al. (2013). "Determining the Noninfectious Complications of Indwelling Urethral Catheters." Ann Intern Med. 159(6): 401-10

Clifton, M., Kralovic, S. M., Simbartl, L. A., Minor, L., Hasselbeck, R., Martin, T., & Roselle, G. A. (2018). Achieving balance between implementing effective infection prevention and control practices and maintaining a home-like setting in US Department of Veterans Affairs nursing homes. American journal of infection control, 46(11), 1307-1310.

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

This is not applicable.

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

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

This is not applicable.

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

What is the source of the systematic review of the body of evidence that supports the performance measure? A systematic review is a scientific investigation that focuses on a specific question and uses explicit, prespecified scientific methods to identify, select, assess, and summarize the findings of similar but

separate studies. It may include a quantitative synthesis (meta-analysis), depending on the available data. (IOM)

☐ Clinical Practice Guideline recommendation (with evidence review)

☐ US Preventive Services Task Force Recommendation

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

☐ Other

1. Guideline for Prevention of Catheter-Associated Urinary Tract Infections 2009. Centers for Disease Control and Prevention. Updated February 2017. https://www.cdc.gov/infectioncontrol/guidelines/CAUTI/index.html#anchor_1552413639

FROM CDC: Summary of Recommendations

Table 1.
Modified Healthcare Infection Control Practices Advisory Committee Categorization (HICPAC) Scheme* for Recommendations

Rank	Description
Category IA	A strong recommendation supported by high to moderate quality† evidence suggesting net clinical benefits or harms
Category IB	A strong recommendation supported by low quality evidence suggesting net clinical benefits or harms or an accepted practice (e.g., aseptic technique) supported by low to very low quality evidence
Category IC	A strong recommendation required by state or federal regulation.
Category II	A weak recommendation supported by any quality evidence suggesting a tradeoff between clinical benefits and harms
No recommendation/unresolved issue	Unresolved issue for which there is low to very low quality evidence with uncertain tradeoffs between benefits and harms

I. Appropriate Urinary Catheter Use

A. Insert catheters only for appropriate indications (see ***Table 1*** for guidance), and leave in place only as long as needed. (Category IB)
   1. Minimize urinary catheter use and duration of use in all patients, particularly those at higher risk for CAUTI or mortality from catheterization such as women, the elderly, and patients with impaired immunity**. (Category IB)**
   2. Avoid use of urinary catheters in patients and nursing home residents for management of incontinence. **(Category IB)**
      a. Further research is needed on periodic (e.g., nighttime) use of external catheters (e.g., condom catheters) in incontinent patients or residents and the use of catheters to prevent skin breakdown. **(No recommendation/unresolved issue)**
   3. Use urinary catheters in operative patients only as necessary, rather than routinely. **(Category IB)**
   4. For operative patients who have an indication for an indwelling catheter, remove the catheter as soon as possible postoperatively, preferably within 24 hours, unless there are appropriate indications for continued use. **(Category IB)**

B. Consider using alternatives to indwelling urethral catheterization in selected patients when appropriate
   1. Consider using external catheters as an alternative to indwelling urethral catheters in cooperative male patients without urinary retention or bladder outlet obstruction. **(Category II)**
   2. Consider alternatives to chronic indwelling catheters, such as intermittent catheterization, in spinal cord injury patients. **(Category II)**
   3. Intermittent catheterization is preferable to indwelling urethral or suprapubic catheters in patients with bladder emptying dysfunction. **(Category II)**
   4. Consider intermittent catheterization in children with myelomeningocele and neurogenic bladder to reduce the risk of urinary tract deterioration. **(Category II)**
   5. Further research is needed on the benefit of using a urethral stent as an alternative to an indwelling catheter in selected patients with bladder outlet obstruction. **(No recommendation/unresolved issue)**
   6. Further research is needed on the risks and benefits of suprapubic catheters as an alternative to indwelling urethral catheters in selected patients requiring short- or long-term catheterization, particularly with respect to complications related to catheter insertion or the catheter site. **(No recommendation/unresolved issue)**

II. Proper Techniques for Urinary Catheter Insertion

A. Perform hand hygiene immediately before and after insertion or any manipulation of the catheter device or site **(Category IB)**
B. Ensure that only properly trained persons (e.g., hospital personnel, family members, or patients themselves) who know the correct technique of aseptic catheter insertion and maintenance are given this responsibility **(Category IB)**
C. In the acute care hospital setting, insert urinary catheters using aseptic technique and sterile equipment **(Category IB)**
    1. Use sterile gloves, drape, sponges, an appropriate antiseptic or sterile solution for periurethral cleaning, and a single-use packet of lubricant jelly for insertion **(Category IB)**
    2. Routine use of antiseptic lubricants is not necessary **(Category II)**
    3. Further research is needed on the use of antiseptic solutions vs. sterile water or saline for periurethral cleaning prior to catheter insertion **(No recommendation/unresolved issue)**
D. In the non-acute care setting, clean (i.e., non-sterile) technique for intermittent catheterization is an acceptable and more practical alternative to sterile technique for patients requiring chronic intermittent catheterization **(Category IA)**
    1. Further research is needed on optimal cleaning and storage methods for catheters used for clean intermittent catheterization **(No recommendation/unresolved issue)**
E. Properly secure indwelling catheters after insertion to prevent movement and urethral traction **(Category IB)**
F. Unless otherwise clinically indicated, consider using the smallest bore catheter possible, consistent with good drainage, to minimize bladder neck and urethral trauma **(Category II)**
G. If intermittent catheterization is used, perform it at regular intervals to prevent bladder overdistension **(Category IB)**
H. Consider using a portable ultrasound device to assess urine volume in patients undergoing intermittent catheterization to assess urine volume and reduce unnecessary catheter insertions **(Category II)**
    1. If ultrasound bladder scanners are used, ensure that indications for use are clearly stated, nursing staff are trained in their use, and equipment is adequately cleaned and disinfected in between patients **(Category IB)**

III. Proper Techniques for Urinary Catheter Maintenance

A. Following aseptic insertion of the urinary catheter, maintain a closed drainage system **(Category IB)**
    1. If breaks in aseptic technique, disconnection, or leakage occur, replace the catheter and collecting system using aseptic technique and sterile equipment **(Category IB)**
    2. Consider using urinary catheter systems with preconnected, sealed catheter-tubing junctions **(Category II)**
B. Maintain unobstructed urine flow **(Category IB)**
    1. Keep the catheter and collecting tube free from kinking **(Category IB)**
    2. Keep the collecting bag below the level of the bladder at all times. Do not rest the bag on the floor **(Category IB)**

3. Emptying the collecting bag regularly using a separate, clean collecting container for each patient; avoid splashing, and prevent contact of the drainage spigot with the nonsterile collecting container **(Category IB)**

C. Use Standard Precautions, including the use of gloves and gown as appropriate, during any manipulation of the catheter or collecting system **(Category IB)**

D. Complex urinary drainage systems (utilizing mechanisms for reducing bacterial entry such as antiseptic-release cartridges in the drain port) are not necessary for routine use **(Category II)**

E. Changing indwelling catheters or drainage bags at routine, fixed intervals is not recommended. Rather, it is suggested to change catheters and drainage bags based on clinical indications such as infection, obstruction, or when the closed system is compromised **(Category II)**

F. Unless clinical indications exist (e.g., in patients with bacteriuria upon catheter removal post urologic surgery), do not use systemic antimicrobials routinely to prevent CAUTI in patients requiring either short or long-term catheterization **(Category IB)**

1. Further research is needed on the use of urinary antiseptics (e.g., methenamine) to prevent UTI in patients requiring short-term catheterization **(Category IB)**

G. Do not clean the periurethral area with antiseptics to prevent CAUTI while the catheter is in place. Routine hygiene (e.g., cleansing of the meatal surface during daily bathing or showering) is appropriate **(Category IB)**

H. Unless obstruction is anticipated (e.g., as might occur with bleeding after prostatic or bladder surgery) bladder irrigation is not recommended **(Category II)**

1. If obstruction is anticipated, closed continuous irrigation is suggested to prevent obstruction **(Category II)**

I. Routine irrigation of the bladder with antimicrobials is not recommended **(Category II)**

J. Routine instillation of antiseptic or antimicrobial solutions into urinary drainage bags is not recommended **(Category II)**

K. Clamping indwelling catheters prior to removal is not necessary **(Category II)**

L. Further research is needed on the use of bacterial interference (i.e., bladder inoculation with a nonpathogenic bacterial strain) to prevent UTI in patients requiring chronic urinary catheterization **(No recommendation/unresolved issue)**

2. Urinary Incontinence Clinical Practice Guideline. Updated 2015. https://paltc.org/product-store/urinary-incontinence-cpg

FROM AMDA: Clinical Practice Guideline: Urinary Incontinence

Urinary incontinence is one of the most common conditions among residents of post-acute and long term care facilities, affecting about 55% of residents. The prevalence of urinary incontinence increases with age. Women are affected more than men. Although urinary incontinence is increasingly prevalent with age, it is not a normal part of aging.

Urinary incontinence can adversely affect patients' dignity and can contribute to depression, embarrassment, and social isolation. The annual cost of managing urinary incontinence in post-acute and long term care facilities is estimated at $5.5 billion.

Outcomes that may be expected from the implementation of this clinical practice guideline include the following:

- Better identification of individuals who have a reversible urinary incontinence problem.
- More individualized approaches to urinary incontinence management.
- More effective targeting of staff resources to urinary incontinence management.
- Minimization of inappropriate use of diapers and catheters.
- Reduction in significant complications of urinary incontinence and urinary catheters.

3. AIPC-SHEA guideline: Infection prevention and control in the long-term care facility. Contains information on indwelling catheter use. September 2008. http://www.apic.org/Resource_/TinyMceFileManager/Practice_Guidance/id_APIC-SHEA_GuidelineforICinLTCFs.pdf

   While the use of urinary catheters in long term care facility residents has decreased in recent years, utilization remains around 5 percent. In long term care residents, the use of invasive devices (e.g., central venous catheters, mechanical ventilators, enteral feeding tubes) increases the likelihood of a device-associated infection (p. 506).

   Guidelines for prevention of catheter-associated UTIs in hospitalized patients are generally applicable to catheterized residents in long-term care facilities. Recommended measures include limiting use of catheters, insertion of catheters aseptically by trained personnel, use of as small diameter a catheter as possible, handwashing before and after catheter manipulation, maintenance of a closed catheter system, avoiding irrigation unless the catheter is obstructed, keeping the collecting bag below the bladder, and maintaining good hydration in residents. Urinary catheters coated with antimicrobial materials have the potential to decrease UTIs but have not been studied in the long-term care setting. For some residents with impaired voiding, intermittent catheterization is an option, and clean technique is as safe as sterile technique. External catheters are also a risk factor for UTIs in male residents but are significantly more comfortable and associated with fewer adverse effects, including symptomatic urinary infection, than indwelling catheters.

4. Diagnosis, Prevention, and Treatment of Catheter-Associated Urinary Tract Infections in Adults: 2009 International Clinical Practice Guidelines from the Infectious Disease Society of America. March 2010. https://academic.oup.com/cid/article/50/5/625/324341

Table 2.
Strength of Recommendation and Quality of Evidence

Category/grade	Definition
A	Good evidence to support a recommendation for or against use.
B	Moderate evidence to support a recommendation for or against use.
C	Poor evidence to support a recommendation for or against use.
Quality of evidence	**Definition**
I	Evidence from >1 properly randomized, controlled trial.
II	Evidence from >1 well-designed clinical trial, without randomization; from cohort or case-controlled analytic studies (preferably from >1 center); from multiple time-series; or from dramatic results from uncontrolled experiments.
III	Evidence from opinions of respected authorities, based on clinical experience, descriptive studies, or reports of expert committees.

Reduction of Inappropriate Urinary Catheter Insertion and Duration

Limiting Unnecessary Catheterization:

1. Indwelling catheters should be placed only when they are indicated **(A-III)**
   i. Indwelling urinary catheters should not be used for the management of urinary incontinence **(A-III)**. In exceptional cases, when all other approaches to management of incontinence have not been effective, it may be considered at patient request.
2. Institutions should develop a list of appropriate indications for inserting indwelling urinary catheters, educate staff about such indications, and periodically assess adherence to the institution-specific guidelines **(A-III)**.
3. Institutions should require a physician's order in the chart before an indwelling catheter is placed **(A-III)**.
4. Institutions should consider use of portable bladder scanners to determine whether catheterization is necessary for postoperative patients **(B-II)**.

Discontinuation of Catheter:

5. Indwelling catheters should be removed as soon as they are no longer required to reduce the risk of CA-bacteriuria **(A-I)** and CA-UTI **(A-II)**.
6. Institutions should consider nurse-based or electronic physician reminder systems to reduce inappropriate urinary catheterization **(A-II)** and CA-UTI **(A-II)**.
7. Institutions should consider automatic stop-orders to reduce inappropriate urinary catheterization **(B-I)**.

Strategies to Consider Prior to Catheter Insertion

Infection Prevention:

8. Hospitals and LTCFs should develop, maintain, and promulgate policies and procedures for recommended catheter insertion indications, insertion and maintenance techniques, discontinuation strategies, and replacement indications **(A-III)**.
    i. Strategies should include education and training of staff relevant to these policies and procedures **(A-III)**.
9. Institutions may consider feedback of CA-bacteriuria rates to nurses and physicians on a regular basis to reduce the risk of CA-bacteriuria **(C-II)**.
    i. Data are insufficient to make a recommendation as to whether such an intervention might reduce the risk of CA-UTI.
10. Data are insufficient to make a recommendation as to whether institutions should place patients with indwelling urinary catheters in different rooms from other patients who have indwelling urinary catheters or other invasive devices to reduce the risk of CA-bacteriuria or CA-UTI.

Alternatives to Indwelling Urethral Catheterization:

11. In men for whom a urinary catheter is indicated and who have minimal postvoid residual urine, condom catheterization should be considered as an alternative to short-term **(A-II)** and long-term **(B-II)** indwelling catheterization to reduce CA-bacteriuria in those who are not cognitively impaired.
    i. Data are insufficient to make a recommendation as to whether condom catheterization is preferable to short-term or long-term indwelling urethral catheterization for reduction of CA-UTI.
    ii. Data are insufficient to make a recommendation as to whether condom catheterization is preferable to short-term or long-term indwelling urethral catheterization for reduction of CA-bacteriuria in those who are cognitively impaired.
12. Intermittent catheterization should be considered as an alternative to short-term **(C-I)** or long-term **(A-III)** indwelling urethral catheterization to reduce CA-bacteriuria and an alternative to short-term **(C-III)** or long-term **(A-III)** indwelling urethral catheterization to reduce CA-UTI.
13. Suprapubic catheterization may be considered as an alternative to short-term indwelling urethral catheterization to reduce CA-bacteriuria **(B-I)** and CA-UTI **(C-III)**.
    i. Data are insufficient to make a recommendation as to whether suprapubic catheterization is preferable to long-term indwelling urethral catheterization for reduction of CA-bacteriuria or CA-UTI.
    ii. Data are insufficient to make a recommendation as to whether intermittent catheterization is preferable to suprapubic catheterization for reduction of CA-bacteriuria or CA-UTI.

1a.4 OTHER SOURCE OF EVIDENCE

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

This is not applicable.

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

This is not applicable.

1b. Performance Gap

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

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

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

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

Significance to residents:

Catheter use is important to address because of its impact on residents. A study by Hu et al. (2014) found that inappropriate catheter use was associated with corollary outcomes of longer hospital stays, increased rate of catheterization at the time of discharge, higher incidence of UTI, and decrease in ADL function (Hu et al., 2014). Hollingsworth et al. (2013) identified additional corollary outcomes of indwelling catheterization. In a meta-analysis of 37 studies involving 2,868 adults, the authors found increased risk of urine leakage in both short- (10.6%) and long-term (52.1%) catheterization. The authors also identified a high risk of urethral strictures (3.4%) among individuals with short-term catheters. Among individuals with spinal cord injuries, gross hematuria (13.5%) and bladder cancer (1.0%) were associated with indwelling catheter use. Although this meta-analysis includes a range of patient types, not limited only to nursing home residents, the findings show that limiting catheter use may be important for preventing a host of serious health issues (Hollingsworth et al., 2013).

Gaps in performance in nursing homes:

Several studies have found that structural characteristics of nursing homes can have an impact on catheter use. Castle and Anderson (2011) found that improvements of structural characteristics, such as increasing RN staffing levels and decreasing RN turnover, significantly decreased catheter use, thus improving resident outcomes. Nursing homes with more educated staff, including a Director of Nurses with at least a bachelor's degree, were found to have lower catheterization rates than those with less educated staff (Damkoehler, 2014). Furthermore, providers with more committed staff through a stronger culture, more experienced staff, and proper staff to resident/patient ratios have also been presented as characteristics that may be linked to adherence with infection control guidelines/practices and appropriate catheter use (Flodgren et al., 2013).

Clifton et al. (2018) studied a quality improvement initiative across Veterans Administration (VA) hospitals and found that use of best practices for catheter insertion, indication, and assessments were correlated with an improved catheter-associated urinary tract infection (CAUTI) outcome (Clifton et al., 2018). Additional processes such as a thorough and informed assessment of the resident, and evaluation of the medical need for a catheter, can sometimes decrease or prevent the use of catheters and the risks associated with their use.

Gaps in performance among specific groups of nursing home residents:

A few studies found an empirical association between social risk factors that could be measured by items available in the MDS 3.0 and catheter use, but did not offer a conceptual basis for understanding how the inherent characteristics of the social risk factor (gender, age, Medicaid coverage, and race/ethnicity) would affect catheter use.

One older study (Rogers et al., 2008) discusses a conceptual basis for using gender as a clinical risk adjuster. However, it is possible that the association between gender and catheter use is reflecting other gender-specific conditions for which facilities should be held accountable when providing care. Similarly, studies looked at the association between age and risk of catheterization and found that older adults are more likely to be catheterized due to a higher likelihood of experiencing adverse health outcomes when being transferred from the acute care setting to SNFs (Burke et al., 2018).

We examined the percentage of long-stay residents with each social risk factor identified in the literature as having an empirical association with catheter use, compared to those without that social risk factor, and used Chi-Squared tests to determine whether these differences were statistically significant.

While all differences were statistically significant and absolute percentages are small, the differences across most subpopulations are relatively large. Among residents who are eligible for Medicaid, 2.0% have/had a catheter inserted and left in their bladder and, among those ineligible for Medicaid, 3.0% have/had a catheter inserted (?2(1) = 472.3, $p < 0.001$). For residents aged 85 years or older, 1.8% have/had a catheter inserted and left in their bladder, compared to 2.4% of younger residents (?2(1) = 474.8, $p < 0.001$). Whereas 3.3% of the male residents have/had a catheter inserted and left in their bladder, 1.6% of the female residents have/had a catheter inserted (?2(1) = 3,200, $p < 0.001$). Alternatively, the difference across race/ethnicity subpopulations is relatively small. Among residents who are non-Hispanic white, 2.1% have/had a catheter inserted and left in their bladder and, among those who are non-white, 2.3% have/had a catheter inserted (?2(1) = 33.7, $p < 0.001$).

Importance to stakeholders:

On May 23, 2019, RTI International convened a web-based technical expert panel (TEP) meeting to obtain expert input on future directions for measure development and maintenance of quality measures for nursing homes based on the Minimum Data Set 3.0. In the pre-TEP survey, 6 out of 10 TEP members rated this measure as “very important” (scoring it a 4 or 5 out of a scale from 1–5), according to the following criteria : is an established priority area (National Quality Strategy); addresses a demonstrated high-impact aspect of health care (e.g., affects large numbers); has external evidence of importance, such as consensus standards; and has evidence of disparities for the quality domain. The majority of TEP members explicitly affirmed the face validity of NQF #0686.

Most TEP members viewed the measure as important because it kept attention on removing catheters and reducing catheter use where possible, given the relationship between catheterization and adverse health outcomes and reduced quality of life. TEP members concurred that measuring catheterization is important and that the measure encourages facilities to focus on continence care. Several TEP members acknowledged that this measure supports quality improvement, with one TEP member noting that this QM encourages facilities to focus on reducing catheterization that is done for convenience or without clinical justification (RTI International, 2019).

Castle, N. G., Anderson, R. A. (2011). Caregiver staffing in nursing homes and their influence on quality of care: using dynamic panel estimation methods. Medical Care, 49(6), 545-552

Clifton, M., Kralovic, S. M., Simbartl, L. A., Minor, L., Hasselbeck, R., Martin, T., & Roselle, G. A. (2018). Achieving balance between implementing effective infection prevention and control practices and maintaining a home-like setting in US Department of Veterans Affairs nursing homes. American journal of infection control, 46(11), 1307-1310.

Damkoehler, G. Using a nurse-led protocol to reduce infections. McKnight’s. Oct. 15, 2014. Available from: http://www.mcknights.com/marketplace/using-a-nurse-led-protocol-to-reduce-infections/article/377448/

Flodgren, G., Conterno, L. O., Mayhew, A., Omar, O., Pereira, C. R., & Shepperd, S. (2013). Interventions to improve professional adherence to guidelines for prevention of device-related infections. Cochrane Database of Systematic Reviews, (3).

Hollingsworth, J. M., et al. (2013). “Determining the Noninfectious Complications of Indwelling Urethral Catheters.” Ann Intern Med. 159(6): 401-10

Hu, F. W., et al. (2014). "Inappropriate use of urinary catheters among hospitalized elderly patients: Clinician awareness is key." Geriatr Gerontol Int. Dec. 11.

RTI International. (2019, June). Technical Expert Panel Summary Report: Maintenance of Nursing Home Quality Measures Prepared under CMS Contract No. HHSM-500-2013-13015I.

RTI analysis of Q3, 2018 MDS 3.0 Data (programming reference: LJC67/LJC09_request_686_31_32_rev.log)

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

Current Measure Performance:

Below we present the distribution of facility-level scores on this quality measure in Quarter 3, 2018. Overall, 1,041,783 long-stay residents in 14,374 nursing homes are included in the analysis. The national facility-level mean score for this measure in Quarter 3, 2018 was 2.2% and the median score was 1.7%, suggesting a slight positive skew. The interquartile range for this measure was 3.4% and 35.2% of facilities had a perfect score of 0.0%. This analysis was restricted to facilities with at least 20 residents in the denominator, the minimum denominator threshold for public reporting.

In Quarter 3, 2018:

k (facilities) 14,374

n (residents) 1,041,783

mean 2.2%

standard deviation (SD) 2.6%

min 0.0%

max 28.5%

Interquartile Range 3.4%

10th percentile 0.0%

20th percentile 0.0%

30th percentile 0.0%

40th percentile 1.1%

50th percentile 1.7%

60th percentile 2.3%

70th percentile 3.0%

80th percentile 3.9%

90th percentile 5.5%

SOURCE: RTI analysis of Q3, 2018 MDS 3.0 data (programming reference: rn27_47\ac362_request_q3132_686.log)

Performance Over Time:

For comparison over time, we also present the distribution of facility-level scores on this quality measure un Quarter 4, 2017. During this time period, 1,033,746 long-stay residents in 14,390 nursing homes are included in the analysis. The national facility-level mean score for this measure in Quarter 4, 2017 was 2.4% and the median score was 1.8%, indicating a slight positive skew. The interquartile range for this measure was 3.6%

and 34.3% of facilities had a perfect score of 0.0%. This analysis was restricted to facilities with at least 20 residents in the denominator, the minimum denominator threshold for public reporting at the time of analysis. Note, the Nursing Home Compare site changed their public reporting restrictions from 30 qualifying residents to 20 qualifying residents for long-stay measures, effective July 2016; when the minimum denominator threshold was larger, fewer facilities had publicly reportable scores.

In Quarter 4, 2017:

k (facilities) 14,390

n (residents) 1,033,746

mean 2.4%

standard deviation (SD) 2.7%

min 0.0%

max 28.2%

Interquartile Range 3.6%

10th percentile 0.0%

20th percentile 0.0%

30th percentile 0.0%

40th percentile 1.2%

50th percentile 1.8%

60th percentile 2.4%

70th percentile 3.2%

80th percentile 4.2%

90th percentile 5.9%

SOURCE: RTI analysis of Q4, 2017 MDS 3.0 data (programming reference: ac04\ac362_request_q2829_686.log)

National facility-level mean and median scores for all available quarters (i.e., Quarter 1, 2011, to Quarter 3, 2018) are presented in the response to 2b1.3. in the Testing Attachment (refer to Figure 1). Since Quarter 1, 2011 the national facility-level mean and median scores have trended steadily downward since the adoption of the MDS 3.0, both decreasing almost monotonically, indicating an overall improvement in performance among providers over time. The mean score for this measure was 4.4% in Quarter 1, 2011, and the median score was 3.7%. In Quarter 3, 2018, the mean and median scores were 2.2% and 1.7%, respectively.

SOURCE: RTI analysis of Q1, 2011 – Q3, 2018 MDS 3.0 data (programming reference: RN27_47\RN18_request_686_add_31_32.log, RN18\RN18_request_686.log)

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

This is not applicable. The data are not estimates based on samples; rather, the data include all long-stay nursing home residents nationally who do not meet exclusion criteria.

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

Current Measure Performance:

We used national Minimum Data Set (MDS) 3.0 data to create the long-stay nursing home resident episode file for Quarter 3, 2018, to examine whether there may be disparities in care for population groups related to this measure. Disparities for certain population groups would indicate gaps in care and opportunities for improvement. In Quarter 3, 2018 there were 15,241 eligible facilities containing 1,118,025 residents eligible for inclusion in the measure (before applying sample size restrictions and without excluding residents that do not have a prior and target assessment); 14,374 facilities (94.3%) containing 1,041,783 residents (93.2%) had sufficient sample sizes (20 or more long-stay residents included in the denominator) to report on this measure after applying all measure exclusion criteria and facility sample size restrictions.

We address the issue of disparities for this measure by examining whether there are differences in catheter use rates for population groups that may reflect disparities in care, such as for population groups with social risk factors. At the resident-level, we examined potential sex (male versus female), race/ethnic (non-Hispanic white versus non-white), insurance status (Medicaid versus non-Medicaid), and age (85 or above versus less than 85) disparities at the resident level by identifying the frequency and proportion of residents across each disparities category who have/had a catheter inserted and left in their bladder.

Referring to the table below, differences in triggering the numerator for NQF #0686 across each disparities category were statistically significant ($p<0.001$).While the absolute percentages of triggering the numerator are small among each resident characteristic, the differences between most subpopulations within a disparities category are relatively large. For residents aged 85 years or older, 1.8% have/had a catheter inserted and left in their bladder, compared to 2.4% of younger residents (?2(1) = 474.8, $p < 0.001$). Whereas 3.3% of the male residents have/had a catheter inserted and left in their bladder, 1.6% of the female residents have/had a catheter inserted (?2(1) = 3,200, $p < 0.001$).Among residents who are eligible for Medicaid, 2.0% have/had a catheter inserted and left in their bladder and, among those ineligible for Medicaid, 3.0% have/had a catheter inserted (?2(1) = 472.3, $p < 0.001$). Alternatively, the difference across race/ethnicity subpopulations is relatively small. Among residents who are non-Hispanic white, 2.1% have/had a catheter inserted and left in their bladder and, among residents who are non-white, 2.3% have/had a catheter inserted (?2(1) = 33.7, $p < 0.001$).

The table below presents frequencies and percentages by disparity group. Residents who were younger, male, non-Medicaid eligible, and identified as non-white were more likely to have a catheter inserted and left in their bladder when compared against their counterparts. For males, this finding was expected.

NQF #0686 Percent of Residents Who Have/Had a Catheter Inserted and Left in Their Bladder (Long Stay) by Social Risk Factors

Resident characteristic (% of all long stay residents)	Frequency of residents who have/had a catheter inserted and left in their bladder	Percentage of residents who have/had a catheter inserted and left in their bladder	Pearson chi2 P-value
Age (1,033,727)			
= 85 (39.93%)	7,383	1.8%	<0.001
< 85 (60.07%)	15,066	2.4%	
Gender (1,033,727)			
Male (34.10%)	11,601	3.3%	<0.001
Female (65.90%)	10,848	1.6%	
Medicaid (947,084)			
Medicaid (88.31%)	16,378	2.0%	<0.001
Non-Medicaid (11.69%)	3,265	3.0%	
Race/ethnicity (1,033,727)			

Non-Hispanic white (73.67%) 16,160 2.1% <0.001

Non-white (26.33%) 6,289 2.3%

Source: RTI analysis of Q3, 2018 MDS 3.0 Data (programming reference: LJC67/LJC09_request_686_31_32_rev.log)

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

This is not applicable; performance data provided in 1b.4.

2. Reliability and Validity—Scientific Acceptability of Measure Properties

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

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

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

Genitourinary (GU)

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

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

Elderly, Populations at Risk, Populations at Risk : Individuals with multiple chronic conditions

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

http://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/NursingHomeQualityInits/NHQIQualityMeasures.html; please see "MDS 3.0 QM User's Manual" in the "User's Manuals" zipped folder in the Downloads section at the bottom of the page

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

This is not an eMeasure **Attachment:**

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

No data dictionary **Attachment:**

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

No, this is not an instrument-based measure **Attachment:**

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

Not an instrument-based measure

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

No

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

This is not applicable; no changes have been made to the measure specifications since the last submission.

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

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

The numerator is the number of long-stay nursing home residents in the denominator sample with an episode during the selected quarter with a target assessment that indicates the use of indwelling catheters within the last seven days.

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

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

The numerator is the number of long-stay nursing home residents in the denominator sample with an episode during the selected quarter with a target assessment that indicates the use of indwelling catheters within the last seven days(H0100A = [1]). For every calendar quarter (3-month period), the Centers for Medicare & Medicaid Services (CMS) select episodes for long-stay residents during that quarter from each nursing home and use the target assessment from that episode 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.

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 home care. Residents who return to the nursing home following a hospital discharge will not have their cumulative days in facility reset to zero.

The target population includes all long-stay residents with a target assessment (assessments may be an OBRA admission, quarterly, annual or significant change/correction assessment (A0310A = [01, 02, 03, 04, 05, 06]); or PPS 5-, 14-, 30-, 60-, 90-day assessments (A0310B = [01, 02, 03, 04, 05]); or discharge assessment with or without anticipated return (A0310F = [10, 11])), except those with exclusions (specified in S.8 and S.9).

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 website and are weighted on an average of four target periods.

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

The denominator includes all long-stay residents in the nursing home who have an episode during the selected quarter with a qualifying target assessment (OBRA, PPS, or discharge assessment) and who do not meet the exclusion criteria.

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

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

Residents are counted in the denominator if they are long-stay residents, defined as residents who have had 101 or more cumulative days of nursing home care. Residents who return to the nursing home after a hospital discharge will not have their cumulative days in facility reset to zero. The target population includes all long-stay residents with a target assessment during the selected quarter (assessments may be an OBRA admission, quarterly, annual or significant change/correction assessment (A0310A = 01, 02, 03, 04, 05, 06); or PPS 5-, 14-, 30-, 60-, 90-day assessments (A0310B = 01, 02, 03, 04, 05); or discharge assessment with or without anticipated return (A0310F = 10, 11)), except for those who meet the exclusion criteria (specified in S.8 and S.9).

A description of the time period for the data included in this measure is provided in S.5 above.

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

The denominator exclusion criteria for this quality measure are as follows: 1) The target assessment is an admission assessment, a PPS 5-day assessment or a PPS readmission/return assessment; 2) The target assessment indicates that indwelling catheter status is missing; 3) The target assessment indicates neurogenic bladder or neurogenic bladder status is missing; or 4) The target assessment indicates obstructive uropathy or obstructive uropathy status is missing.

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

If the target assessment is an admission assessment (A0310A = [01]), PPS 5-day assessment (A0310B = [01]) or PPS

readmission/return anticipated assessment (A0310B = [06]), the resident is excluded.

A resident is also excluded if any of the following conditions are true:

1) Target assessment indicates that indwelling catheter status is missing (H0100A = [-]).

2) Target assessment indicates neurogenic bladder (I1550 = [1]) or neurogenic bladder status is missing (I1550 = [-]).

3) Target assessment indicates obstructive uropathy (I1650 = [1]) or obstructive uropathy status is missing (I1650 = [-]).

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.

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

This is not applicable; this measure is not stratified.

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

Statistical risk model

If other:

S.12. Type of score:

Rate/proportion

If other:

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

Better quality = Lower score

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

This measure is risk-adjusted for bowel incontinence and pressure ulcers at Stage II, III, or IV using a logistic regression. The measure is calculated as follows:

Step 1: Identify the total number of long-stay residents who do not meet the exclusion criteria, with a selected target assessment (OBRA, PPS, or discharge) during the quarter.

Step 2: Calculate the facility-level observed score (steps 2a through 2b below).

Step 2a: Starting with the set of residents identified in Step 1, determine the total number of long-stay residents with a selected target assessment that meets the numerator inclusion criteria (H0100A = [1]).

Step 2b: Calculate the facility observed score by dividing the results of step 2a by the results of step 1

Step 3: Calculate the national observed score by averaging the scores derived in step 2b across all facilities.

Step 4: Calculate the expected resident score for each resident (steps 4a and 4b below)

Step 4a: Assign covariate values, either '0' for covariate condition not present or '1' for covariate condition present, for the residents included in the denominator for each of the two covariates (bowel incontinence and presence of pressure ulcers) based on the resident's prior assessment and run the logistic regression model.

Specifically, the covariates are calculated as follows:

For the variable identifying frequent bowel incontinence on prior assessment (H0400 = [2, 3]):

1. Covariate = [1] if H0400 = [2, 3];
2. Covariate = [0] if H0400 = [0, 1, 9, -]

For the variable identifying pressure ulcers at stage II, III, or IV on prior assessment:

1. Covariate = [1] if any of the following are true:

a. M0300B1 = [1, 2, 3, 4, 5, 6, 7, 8, 9], or

b. M0300C1 = [1, 2, 3, 4, 5, 6, 7, 8, 9], or

c. M0300D1 = [1, 2, 3, 4, 5, 6, 7, 8, 9].

2. Covariate = [0] if the following is true:

a. M0300B1 = [0, -, ^] and

b. M0300C1 = [0, -, ^] and

c. M0300D1 = [0, -, ^].

*All covariates are missing if no prior assessment is available.

The logistic regression model is of the form:

[Equation 1] QM triggered (yes=1, no=0) = B0 + B1*COVA + B2*COVB

Where:

B0 is the logistic regression constant (B0 =-4.054929),

B1 is the logistic regression coefficient for the first covariate, bowel incontinence (B1 = 0.503225),

COVA is the resident-level score for the first covariate (0 or 1),

B2 is the logistic regression coefficient for the second covariate, pressure ulcers at stage II, III, or IV (B2 = 2.200337, and

COVB is the resident-level score for the second covariate (0 or 1)

Step 4b: Calculate the expected resident score for each resident with the following formula:

[Equation 2] Resident-level expected QM score = 1/ [1+e^-x]

Where e is the base of natural logarithms and x is a linear combination of the constant and the logistic regression coefficients times the covariate scores (from Equation [1], above). A covariate score will be 1 if the covariate is triggered for that resident, and 0 if the covariate is not triggered.

Step 5: Calculate the facility-level expected QM score by averaging all resident-level expected scores derived in step 4b.

Step 6. Calculate the facility-level adjusted score based on the:

• facility-level observed QM score (step 2b),

• facility-level expected QM score (step 5), and

• national average observed QM score (step 3).

The calculation of the adjusted score uses the following equation:

[Equation 3] Adj = 1/ [1 + e^ -y]

where

Adj is the facility-level adjusted QM score, and

y = (Ln(Obs/(1–Obs) - Ln(Exp/(1–Exp) + Ln(Nat/(1–Nat))

Obs is the facility-level observed QM rate,

Exp is the facility-level expected QM rate,

Nat is the national observed QM rate (Nat = 0.028926), and

Ln indicates a natural logarithm.

e is the base of natural logarithms

RTI International. (2019). Analysis of Q3, 2018 MDS 3.0 data (programming reference: rn27_47\LJC10_request_q2829_686.log)

Reference: The Centers for Medicare & Medicaid Services (CMS) (January 2019). MDS 3.0 Quality Measures User's Manual. RTI International, Waltham, MA. Accessed at:

http://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/NursingHomeQualityInits/NHQIQualityMeasures.html; please see "MDS 3.0 QM User's Manual" in the "User's Manuals" zipped folder in the Downloads section at the bottom of the page.

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

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

This is not applicable. The data are not estimates based on samples; rather, the data include all long-stay nursing home residents nationally who do not meet exclusion criteria.

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

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

This is not applicable; this measure is not based on survey/patient-reported data.

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

If other, please describe in S.18.

Assessment Data

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

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

The data source is the Minimum Data Set (MDS) 3.0, and the collection instrument is the Resident Assessment Instrument (RAI) version 1.15.

For MDS 3.0 item sets used to calculate the quality measure, refer to: http://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/NursingHomeQualityInits/NHQIMDS30TechnicalInformation.html

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

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

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

Facility

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

Post-Acute Care

If other:

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

This is not applicable; this is not a composite performance measure.

2. Validity – See attached Measure Testing Submission Form

0686_MeasureTesting_MSF5.0_Data.doc,NQF_0686_Catheter_Testing_Form_Final_7-22-19-636997438877755033.docx

2.1 For maintenance of endorsement

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

Yes

2.2 For maintenance of endorsement

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

Yes

2.3 For maintenance of endorsement

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

Yes - Updated information is included

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

NATIONAL QUALITY FORUM—Measure Testing (subcriteria 2a2, 2b1-2b6)

Measure Number (*if previously endorsed*)**:** 0686
Measure Title: Percent of Residents Who Have/Had a Catheter Inserted and Left in Their Bladder (Long Stay)
Date of Submission: 8/1/2019

Type of Measure:

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

Instructions

- **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.**
- **For all measures, sections 1, 2a2, 2b1, 2b2, and 2b4 must be completed.**
- **For outcome and resource use measures, section 2b3 also must be completed.**
- **If specified for multiple data sources/sets of specificaitons (e.g., claims and EHRs), section 2b5 also must be completed.**
- **Respond to all questions as instructed with answers immediately following the question. All information on testing to demonstrate meeting the subcriteria for reliability (2a2) and validity (2b1-2b6) must be in this form. An appendix for *supplemental* materials may be submitted, but there is no guarantee it will be reviewed.**
- **If you are unable to check a box, please highlight or shade the box for your response.**
- **Maximum of 25 pages (*incuding questions/instructions;* minimum font size 11 pt; do not change margins). *Contact NQF staff if more pages are needed.***
- **Contact NQF staff regarding questions. Check for resources at 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 version 7.1 of the Measure Testing Attachment.**

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.

2a2. Reliability testing [10] 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 [11] 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; [12]

AND

If patient preference (e.g., informed decisionmaking) 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). [13]

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 nonresponders) and how the specified handling of missing data minimizes bias.

Notes

10. 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).

11. 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.

12. 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.

13. Patient preference is not a clinical exception to eligibility and can be influenced by provider interventions.

14. Risk factors that influence outcomes should not be specified as exclusions.

15. 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.

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

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

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

Measure Specified to Use Data From: (***must be consistent with data sources entered in S.17***)	**Measure Tested with Data From:**
☐ abstracted from paper record	☐ abstracted from paper record
☐ claims	☐ claims
☐ registry	☐ registry
☐ abstracted from electronic health record	☐ abstracted from electronic health record
☐ eMeasure (HQMF) implemented in EHRs	☐ eMeasure (HQMF) implemented in EHRs
☒ other: Nursing Home Minimum Data Set (MDS) 3.0 v1.15.0	☒ other: Nursing Home Minimum Data Set (MDS) 3.0 v1.15.0

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

The data set used for testing was the Nursing Home Minimum Data Set (MDS) 3.0 v1.15.0.

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

Two studies were used in the testing of this measure; they are described in greater detail below.

1. RAND Corporation – Development and validation of a revised nursing home assessment tool: MDS 3.0; August 2006 to February 2007 (Saliba & Buchanan, 2008).

2. RTI International – Analysis of MDS 3.0 data: Quarter 2, 2018 and Quarter 3, 2018.
   a. Trend analysis done for Quarter 1, 2011 – Quarter 3, 2018 in Section 2b1 (RTI International, 2019).

RTI International: RTI analysis of MDS 3.0 data for Quarter 2, 2018 and Quarter 3, 2018.

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 http://www.cms.hhs.gov/NursingHomeQualityInits/Downloads/MDS30FinalReport.pdf.

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

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

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

1. 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. This study tested item-level reliability and validity of the items used to identify continence by comparing item-level coding among gold-standard nurses (349 cases) and comparing item-level coding between gold-standard nurse and participating facility staff (900 cases). Additionally, this study sought feedback from providers on the items used to identify urinary continence and toileting programs with the aim of improving the assessment process and reporting of residents' continence status to reduce confusion associated with the continence category and to improve incontinence management (Saliba & Buchanan, 2008).

2. RTI facility-level analyses of MDS 3.0 data sample included all facilities with sufficient sample size ($n \geq 20$ residents) to publicly report this measure in Quarter 3, 2018 ($k = 14{,}374$ facilities), unless otherwise noted (RTI International, 2019).[1]

RTI International (2019). RTI analysis of MDS 3.0 data for Quarter 3, 2018 (programming references: rn27_47\ac356_request_q3132_686.log)

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 http://www.cms.hhs.gov/NursingHomeQualityInits/Downloads/MDS30FinalReport.pdf.

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

[1] To analyze the change in facility performance from one quarter to the next, MDS 3.0 data from Quarter 2, 2018 and Quarter 3, 2018 were used. For these analyses, the sample includes facilities that had a reportable score (minimum denominator ≥ 20 residents) for the measure in both quarters.

1. The RAND Development and Validation of MDS 3.0 study sample included 3,822 residents from community nursing homes and 764 residents from 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 http://www.cms.hhs.gov/NursingHomeQualityInits/Downloads/MDS30FinalReport.pdf.

2. The sample for the RTI analysis of MDS 3.0 includes all long-stay residents counted in the denominator for this measure, after measure exclusion criteria were applied, in facilities with a sufficient sample size ($n \geq 20$, $k = 14{,}374$) to report this measure ($n = 1{,}041{,}783$) in Quarter 3, 2018 (RTI International, 2019).

Table 1a below presents the characteristics of long-stay residents who are counted in the denominator for this measure in Quarter 3, 2018, after measure exclusion criteria were applied but *before* applying facility sample size restrictions and without excluding those residents without both a prior and target assessment (n = 1,118,025); the n for each resident characteristic varies due to the proportion of missing data for that characteristic. Although most analyses include only facilities *after* applying facility sample size restrictions, this table is representative of the pool of candidate residents that may be in the denominator regardless of changes in facility census. ***Table 1b*** offers the characteristics of the residents who are counted in the denominator after applying facility sample size restrictions in Quarter 3, 2018, to clarify the actual description of residents included in the testing and analysis presented for this quality measure as described in **1.6 above**.

Table 1a. Characteristics of Long-Stay Residents Eligible for Inclusion in Analyses of NQF #0686 Percent of Residents Who Have/Had a Catheter Inserted and Left in Their Bladder (Long Stay) (Quarter 3, 2018)

Resident characteristics	Frequency (n)	Total Observations (N)	Percentage (%)
Sex			
Female	720,561	1,118,025	64.4%
Male	397,464	1,118,025	35.6%
Race/Ethnicity			
White Only	824,901	1,118,025	73.8%
Black or African American Only	173,361	1,118,025	15.5%
Hispanic or Latino Only	64,038	1,118,025	5.7%
Asian Only	22,173	1,118,025	2.0%
American Indian/Alaska Native Only	5,225	1,118,025	0.5%
Native Hawaiian or Other Pacific Islander Only	1,646	1,118,025	0.2%
Multi-race	3,399	1,118,025	0.3%
Medicare-Medicaid Dual Eligibility			
Dual-Eligible	808,573	1,118,025	72.3%
Non-Dual	228,616	1,118,025	20.4%
Missing	80,836	1,118,025	7.2%
Age			
<65	183,062	1,118,025	16.4%
65-74	205,513	1,118,025	18.4%
75-84	293,616	1,118,025	26.3%
85+	435,834	1,118,025	39.0%

Resident characteristics	Frequency (n)	Total Observations (N)	Percentage (%)
Diagnoses			
Arthritis	345,211	1,113,759	31.0%
Osteoporosis	148,455	1,113,779	13.3%
Hip Fracture	54,013	1,117,888	4.8%
Other Fracture	87,908	1,117,881	7.9%
Depression	653,120	1,117,888	58.4%
Stroke	185,069	1,117,878	16.6%
Alzheimer's Disease	189,607	1,117,888	17.0%
Non-Alzheimer's Dementia	576,710	1,117,887	51.6%
Malnutrition or at risk for malnutrition	8,2447	1,117,988	7.4%
Cancer	76,269	1,113,742	6.8%
Anemia	406,977	1,117,872	36.4%
Heart Failure	267,707	1,117,893	23.9%
Hypertension	909,057	1,117,881	81.3%
Diabetes Mellitus	409,995	1,117,983	36.7%
Anxiety Disorder	426,683	1,117,977	38.2%
Asthma, Chronic Obstructive Pulmonary Disease, or Chronic Lung Disease	295,384	1,117,891	26.4%

Source: RTI analysis of Q3, 2018 MDS 3.0 data (programming reference: LJC66\LJC66_request_686_31_32.log)

Table 1b. Characteristics of Long-Stay Residents Included in Analyses of NQF #0686 Percent of Residents Who Have/Had a Catheter Inserted and Left in Their Bladder (Long Stay) (Quarter 3, 2018)

Resident characteristics	Frequency (n)	Total Observations (N)	Percentage (%)
Sex			
Female	712,875	1,104,988	64.5%
Male	392,113	1,104,988	35.5%
Race/Ethnicity			
White Only	815,285	1,104,988	73.8%
Black or African American Only	171,366	1,104,988	15.5%
Hispanic or Latino Only	63,282	1,104,988	5.7%
Asian Only	21,957	1,104,988	2.0%
American Indian/Alaska Native Only	5,154	1,104,988	0.5%
Native Hawaiian or Other Pacific Islander Only	1,625	1,104,988	0.2%
Multi-race	3,355	1,104,988	0.3%
Medicare-Medicaid Dual Eligibility			
Dual-Eligible	798,732	1,104,988	72.3%

Resident characteristics	Frequency (n)	Total Observations (N)	Percentage (%)
Non-Dual	226,680	1,104,988	20.5%
Missing	79,576	1,104,988	7.2%
Age			
<65	181,073	1,104,988	16.4%
65-74	202,531	1,104,988	18.3%
75-84	289,992	1,104,988	26.2%
85+	431,392	1,104,988	39.0%
Diagnoses			
Arthritis	341,322	1,100,866	31.0%
Osteoporosis	146,988	1,100,887	13.4%
Hip Fracture	52,928	1,104,852	4.8%
Other Fracture	86,563	1,104,844	7.8%
Depression	645,469	1,104,851	58.4%
Stroke	182,752	1,104,841	16.5%
Alzheimer's Disease	187,904	1,104,851	17.0%
Non-Alzheimer's Dementia	570,895	1,104,850	51.7%
Malnutrition or at risk for malnutrition	81,228	1,104,951	7.4%
Cancer	75,198	1,100,849	6.8%
Anemia	400,970	1,104,835	36.3%
Heart Failure	262,975	1,104,856	23.8%
Hypertension	897,936	1,104,845	81.3%
Diabetes Mellitus	404,006	1,104,946	36.6%
Anxiety Disorder	421,477	1,104,940	38.1%
Asthma, Chronic Obstructive Pulmonary Disease, or Chronic Lung Disease	290,726	1,104,854	26.3%

Source: RTI analysis of Q3, 2018 MDS 3.0 data (programming reference: LJC66\LJC66_request_686_31_32.log)

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

All analyses used the same data as described above in sections 1.5 and 1.6.

Data for Critical Data Elements

RAND reliability analysis of data elements used the same sample as described in sections 1.5 and 1.6 (Saliba & Buchanan, 2008).

Data for Measure Performance Score Testing

RTI analyses used the same data as described in sections 1.5 and 1.6.

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 http://www.cms.hhs.gov/NursingHomeQualityInits/Downloads/MDS30FinalReport.pdf.

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

Analyses are based on resident-level social risk factor variables related to catheterization and available in the MDS 3.0, including race/ethnicity, age, gender, and Medicaid status. We selected these resident-level social risk factors based on literature showing that catheter use and urinary continence care in nursing homes can vary by race/ethnicity, age, gender, and Medicaid status.

2a2. RELIABILITY TESTING

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

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

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

Critical Data Element Reliability

1. The national test of MDS 3.0 items examined the agreement between assessors (reliability). 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. Residents participating in the test were selected to capture a representative sample of short- and long-stay residents. In this national test, the agreement between gold-standard nurses and between gold-standard and facility nurses on bladder and bowel continence, toileting, and catheter items was examined. In addition, with respect to the risk adjustors for this measure, agreement between gold-standard nurses and between gold-standard and facility nurses on items for pressure ulcers were examined. Cohen's kappas 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, where a rating of greater than 0.60 is considered substantial agreement (Landis & Koch, 1977).

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 http://www.cms.hhs.gov/NursingHomeQualityInits/Downloads/MDS30FinalReport.pdf.

Performance Measure Score Reliability

2.a. Signal-to-noise analysis: If a measure is reliable, then true differences in provider performance should explain a substantial proportion of the variance in quality measure scores. A signal-to-noise analysis was performed to determine what proportion of total variance in the measure is attributable to differences among providers. This analysis used logistic regression of the measure numerator triggering for Quarter 3, 2018. We ran a logistic regression analysis with one term (a binary variable

equal to 1 if the measure numerator is triggered and 0 if otherwise; please refer to S.4 and S.5 for more details on the measure numerator specifications) with facility random effects to obtain an estimate of ρ, the proportion of the total variance contributed by the facility-level variance component (i.e., $\rho = \frac{\sigma_v^2}{\sigma_v^2+\sigma_\varepsilon^2}$). The signal-to-noise ratio ρ is a measure of how well a measure can detect differences between facilities. For nursing home quality measures, we typically see values that are 0.1 or lower.

2.b. 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 the nursing facility's quality measure scores on the basis of the two randomly divided halves. When a nursing facility's residents, randomly divided, have similar scores 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, we conducted a split-half reliability analysis on all facilities with 20 or more residents counted in the measure denominator. We used the Pearson Product-Moment Correlation (*r*), Spearman Rank Correlation (*ρ)*, and Intraclass Correlation Coefficient (ICC) to measure the internal reliability.

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

Critical Data Element Reliability

1. For the bladder and bowel continence items, including catheter use, the average kappa for gold-standard nurse to gold-standard nurse agreement was 0.949, and the average kappa for gold-standard nurse to facility nurse agreement was 0.945. For the MDS items for pressure ulcers, average kappa for gold-standard to gold-standard nurse agreement was 0.905, and the average kappa for gold-standard nurse to facility nurse agreement was 0.937. All values for kappa are well above the value of 0.60, which is generally considered substantial agreement (Saliba & Buchanan, 2008).

Performance Measure Score Reliability

2.a. Signal-to-noise: The signal-to-noise ratio for this measure was 0.132 ($p < 0.001$) indicating that 13.2% of the variance in scores for this measure in Quarter 3, 2018 was explained by inter-facility characteristics (including the underlying quality of care in each facility) (RTI International, 2019a). Thus, this measure is somewhat reliable in separating facility characteristics from the noise of population variance.

2.b. Split-half reliability analysis: Correlations above 0.6 are generally considered as evidence of strong reliability (Armitage & Berry, 1994; Bland & Altman, 1986). The split-half correlation for this measure was positive, but the relationship was weak (r = 0.26, ρ = 0.22, p < .001), and the ICC was 0.27 (p < .001), providing limited evidence of internal reliability (RTI International, 2019b). These low correlations were expected due to only a small amount of variation in performance among providers. The national-level distribution indicates one-third of all providers achieved a perfect score of 0.0% (lower scores demonstrate higher quality), giving rise to a positively-skewed distribution of provider performance; perhaps even more significant, the range of non-perfect scores was very small, between 1.7% in the 50th percentile and 5.5% in the 90th percentile (see ***Table 11*** in **Section 2b4.2** where these data are presented. Taking this finding into account, intra-facility splits of the data would result in pairs of zero values (0.0%) being compared for one-third of all facilities. Since correlations are calculated using the covariance of the data, and the individual variances to naturalize the covariance to report a value between -1 and 1, the small amount of variance in performance was expected to yield low correlation coefficients. Further, given the overall high-performance scores among non-perfect providers and the variation in denominator sizes, it's likely that splitting the data would also result in

scores of 0% being compared to non-zero values. This may also result in low correlations because the data may not vary in a systematic or predictable way across the split samples.

Armitage P., & Berry, G. (1994). In: *Statistical Methods in Medical Research,* 3rd edn. Oxford: Blackwell Scientific Publications:312-41.

Bland, J., & Altman, D. (1986). *Statistical methods for assessing agreement between two methods of clinical measurement.* Lancet; i:307-10.

RTI International. (2019a). Analysis of Q3, 2018 MDS 3.0 data (programming reference: rn27_47\ac363_request_q3132_0686.log)

RTI International. (2019b). Analysis of Q3, 2018 MDS 3.0 data (programming reference: rn27_47\IL28_request_q3132_686.log)

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 http://www.cms.hhs.gov/NursingHomeQualityInits/Downloads/MDS30FinalReport.pdf.

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

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.

Performance Measure Score Reliability

RTI's analyses demonstrate that the measure shows some evidence of internal reliability, and the differences in the facilities can explain about thirteen percent of the variance in this measure. Although the signal-to-noise ratio is low, it is moderate in the context of the nursing home quality measures. Additionally, as explained above, there is a strong explanation (i.e., positively skewed distribution of provider performance) why the split-half reliability analysis would yield low explanations.

2b1. VALIDITY TESTING

2b1.1. What level of validity testing was conducted? (*may be one or both levels*)

☐ **Critical data elements** (*data element validity must address ALL critical data elements*)

☒ **Performance measure score**

☒ **Empirical validity testing**

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

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

Performance Measure Score Validity

1.a. Correlation with related quality measures: To assess convergent validity, RTI examined whether a facility's percentile rank on one quality measure in a measure group was correlated with its percentile rank on another quality measure in the same clinically-related group. Specifically, we examined

whether a facility's percentile rank on this measure (NQF# 0686) was correlated with that facility's performance on the related quality measures NQF #0684 (Percent of Residents with a Urinary Tract Infection (Long Stay)) and NQF #0685 Percent of Low Risk Residents Who Lose Control of their Bowel or Bladder (Long Stay)). Although historically low correlations have been observed among nursing home quality measures, we hypothesize that a nursing facility's percentile rank on NQF #0686 and its percentile rank on NQF #0684 should have a positive, but weak, correlation because both measures are concerned with genitourinary care provided to long-stay residents. Similarly, we hypothesize that a nursing facility's percentile rank on NQF #0686 and NQF #0685 should have a positive, but weak, correlation because both measures are concerned with continence-related care provided to long-stay residents.

1.b. Variation by state: We examined whether 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.

1.c. 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, we examined the trend in the national mean and median for this quality measure score between Quarter 1, 2011 and Quarter 3, 2018.

1.d. Stability analysis: We examined the extent to which relative facility rank changed on this quality measure from Quarter 2 to Quarter 3, 2018. We evaluated the percentage of facilities that changed in their percentile ranking (i.e., relative quality measure score) within 1 decile, between 1 and 2 deciles, between 2 and 3 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.

1.e. Confidence interval analysis: We examined proportions of facilities with scores for this measure that are significantly different from the national facility-level mean, stratified by facility denominator size. A valid measure should have a large proportion of facilities with scores significantly different than the mean due to the variation in resident characteristics and conditions among the nursing homes included in the sample. 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. Because this measure is focusing on an undesirable outcome, high-performing facilities should have scores that are significantly below average, and scores of 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.

1.f. Average change in performance across years: We calculated the difference in performance scores for this measure across years to assess how updates to the guidance due to implementation of ICD-10 codes (in October 2015), or updates to guidance in the Long-Term Care facility Resident Assessment Instrument 3.0 User's Manual pertaining to item I2300 – Urinary Tract Infection (UTI) (LAST 30 DAYS) (in October 2017), changed provider scores from year to year. The changes in guidance could compromise the validity of the measure if the variation in the overall or regional facility performance observed on this measure is attributable to this change in guidance or if there are changes in the measure sample. Like the seasonality discussion, this may result in a threat to the measure's validity if providers experience considerable variation or differences in performance across years.

1.g. Face validity: RTI convened a Technical Expert Panel (TEP) on May 23, 2019 to obtain feedback from providers and various stakeholders about the face validity of NQF #0686. TEP members discussed the current measure specifications, including the exclusion criteria of neurogenic bladder and obstructive uropathy, and the risk-adjustment items of pressure ulcers at stages II – IV and frequent bowel incontinence, to determine the face validity of the measure as it is currently specified (RTI International, 2019).

RTI International. (2019, June). Technical Expert Panel Summary Report: Maintenance of Nursing Home Quality Measures Prepared under CMS Contract No. HHSM-500-2013-13015I.

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

Performance Measure Score Validity

1.a. Correlation with related quality measures: Among facilities who could report both measures, RTI calculated the correlation between the facility's percentile rank on NQF #0686 (Percent of Residents Who Have/Had a Catheter Inserted and Left in Their Bladder (Long Stay)) and the facility's percentile rank on NQF #0684 (Percent of Residents with a Urinary Tract Infection (Long Stay)) and found a positive, but weak (ρ = 0.110), and statistically significant ($p < 0.001$) correlation. Among facilities who could report both measures, RTI also calculated the correlation between the facility's percentile rank on NQF #0686 (Percent of Residents Who Have/Had a Catheter Inserted and Left in Their Bladder (Long Stay)) and NQF #0685 (Percent of Low Risk Residents Who Lose Control of their Bowels or Bladder (Long Stay)) and found a negative, but weak (ρ = -0.006) and statistically insignificant relationship.

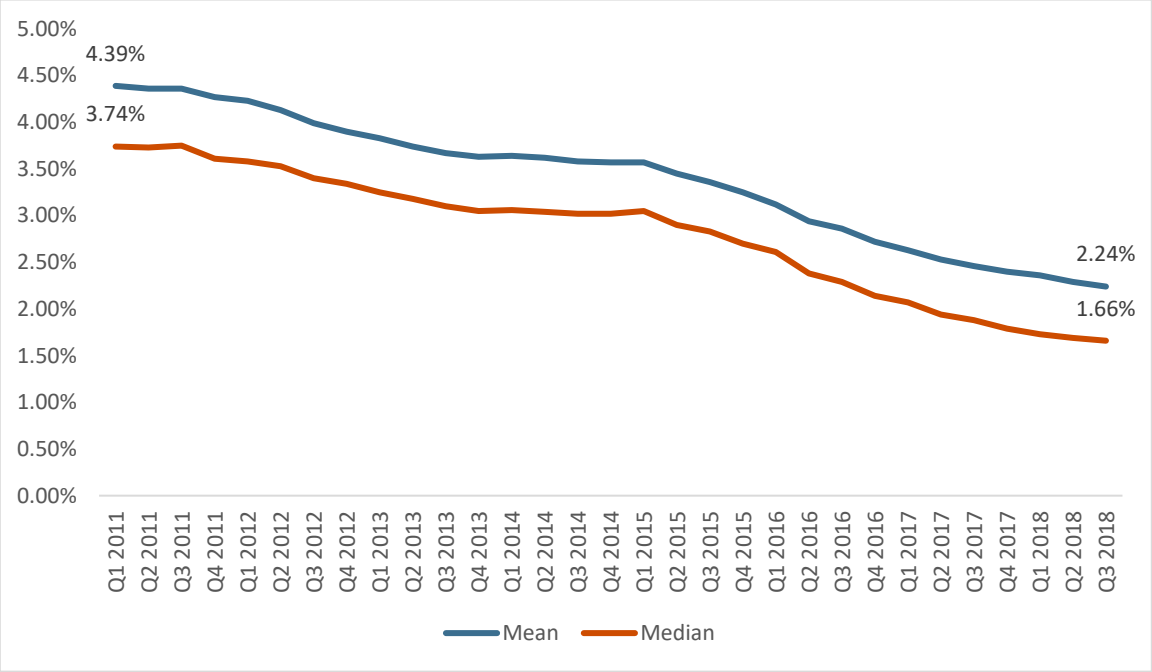
Source: RTI analysis of Q3, 2018 MDS 3.0 data (programming reference: rn27_47\AC359_request_q3132_686.log)

1.b. Variation by State: RTI conducted a one-way analysis of variance (ANOVA) and examined the interquartile range in mean state-level scores across states to assess whether state characteristics were a source of facility measure score variation for NQF #0686. The proportion of variance in this measure explained by the state in which facilities are located is 0.32% ($p < 0.001$). The interquartile range of mean state-level scores is 3.4% (RTI International, 2019b).

Source: RTI analysis of Q3, 2018 MDS 3.0 data (programming reference: rn27_47\AC357_request_q3132_686.log)

1.c. Seasonality: RTI examined the national-level mean and median quality measure scores for each quarter from Quarter 1, 2011, to Quarter 3, 2018. The results are presented in ***Figure 1***. The national-level means and medians have both decreased almost monotonically since Quarter 1 of 2011. These results show no evidence of seasonal variation. Further, the decreasing trend indicates that facilities may have improved their ability to assess residents during this period, for example, by offering individualized treatment and services to achieve or maintain normal elimination function.

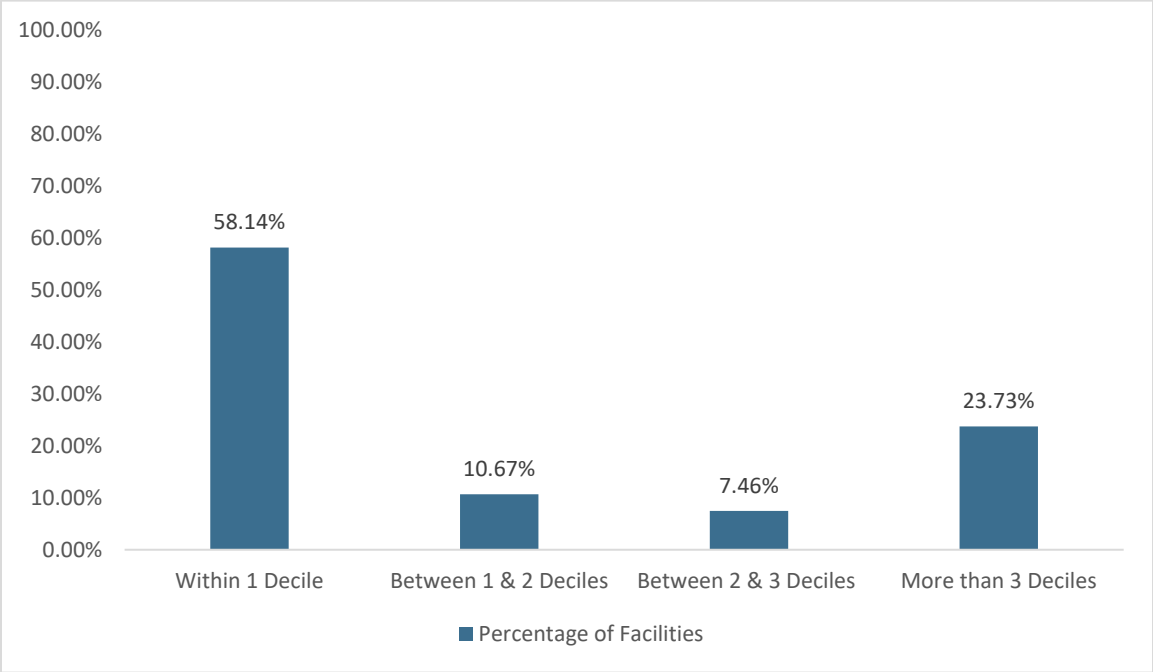
Figure 1. Seasonal (Quarterly) Variation in NQF #0686 Percent of Residents Who Have/Had a Catheter Inserted and Left in Their Bladder (Long Stay)



Source: RTI analysis of Q1, 2011 – Q3, 2018 MDS 3.0 data (programming reference: RN27_47\RN18_request_686_add_31_32.log, RN18\RN18_request_686.log)

1.d. Stability analysis: ***Figure 2*** illustrates the changes in facility rank by quality measure score from Quarter 2, 2018, to Quarter 3, 2018 (k = 14,261). The majority (58.14%) of facilities ranked within the same decile in both quarters. Shifts of more than 3 deciles were less common, occurring for approximately 23.73% of facilities. Thus, both facility scores and relative ranks for this measure are stable from one quarter to the next.

Figure 2. Decile Change in Facility Ranking from Quarter 2, 2018, to Quarter 3, 2018, NQF #0686 Percent of Residents Who Have/Had a Catheter Inserted and Left in Their Bladder (Long Stay)



*Facilities were included in this analysis if they could publicly report a measure score for both Quarter 2 and Quarter 3, 2018 to properly identify the difference in performance across the two quarters.

Source: RTI analysis of Q2 and Q3, 2018 MDS 3.0 data (programming reference: RN27_47\ac358_request_q3132_686.log

1.e. Confidence interval analysis: Another measure of validity is performance relative to the mean: high-performing facilities should have scores that are significantly below-average, and low-performing facilities should be significantly above-average. ***Table 2*** shows the proportions of facilities that scored significantly higher or lower (i.e., different) than the national facility-level mean in Quarter 3, 2018. 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.

In general, there were fewer facilities with quality measure scores that were statistically significantly ($p \leq .05$) higher than the national mean of 2.2% (1.54%) than those with scores that were statistically significantly lower than the national mean (36.55%), indicating that more facilities perform better (lower scores are better) than the national facility-level mean.

The proportions of facilities with scores that are significantly different from the national mean vary as a function of the number of residents included in the denominator for this measure; the percentage of facilities which have scores that are statistically significantly different from the mean generally decreases as number of residents increases.

Overall, just above one-third (38.10%) of facilities were significantly different from the national mean in Quarter 3, 2018, suggesting that there are some meaningful differences in facility-level scores for this measure and providing evidence of validity for NQF #0686.

Table 2. Proportion of Facilities with Scores Significantly Different from the National Facility-Level Mean, Stratified by Facility Denominator Size for NQF #0686 Percent of Residents Who Have/Had a Catheter Inserted and Left in Their Bladder (Long Stay), Quarter 3, 2018

Decile of denominator size in residents	*k*	Number of facilities with 95% confidence interval lower than national mean (%)	Number of facilities with 95% confidence interval higher than national mean (%)	Total number of facilities with scores significantly different from mean (%)
1st Decile (*n* = 20 to 31)	1,520	833 (54.80)	14 (0.92)	847 (55.72)
2nd Decile (*n* = 32 to 39)	1,424	702 (49.30)	13 (0.91)	715 (50.21)
3rd Decile (*n* = 40 to 47)	1,481	643 (43.42)	11 (0.74)	654 (44.16)
4th Decile (*n* = 48 to 54)	1,352	527 (38.98)	17 (1.26)	544 (40.24)
5th Decile (*n* = 55 to 63)	1,565	569 (36.36)	25 (1.60)	594 (37.96)
6th Decile (*n* = 64 to 71)	1,419	451 (31.78)	20 (1.41)	471 (33.19)
7th Decile (*n* = 72 to 80)	1,310	391 (29.85)	22 (1.68)	413 (31.53)
8th Decile (*n* = 81 to 93)	1,469	369 (25.12)	35 (2.38)	404 (27.50)
9th Decile (*n* = 94 to 117)	1,403	357 (25.45)	28 (2.00)	385 (27.44)
10th Decile (*n* = 118 to 654)	1,431	412 (28.79)	37 (2.59)	449 (31.38)
Total (*n* = 20 to 654)	**14,374**	**5,254 (36.55)**	**222 (1.54)**	**5,476 (38.10)**

NOTE: *k* = number of facilities that meet minimum requirements for public reporting this quality measure.
Source: RTI analysis of Q3, 2018 MDS 3.0 data (programming reference: rn27_47\ac364_request_q3132_686.log)

1.f. Average change in performance across years: ***Table 3*** presents the changes in provider performance scores from year to year, using fiscal year (FY) 2013 – 2018. On average, provider scores changed by less than 0.01 percentage points on NQF #0686 between years. Few facilities experienced a change in performance by 0.05 or greater and over 90% of provider scores changed by 0.03 points or less between years. The mean nursing home score change between FY 2017 and FY 2018 (the coding guideline changed at the beginning of FY 2018) was 0.003, which is similar to previous mean facility-score changes between other years where there was no change in clinical coding guidelines (0.001 – 0.005). While impacts of the change that occurred on October 1st, 2015 with the implementation of the ICD-10 codes will continue to be monitored as more data become available, there is no indication that there was a significant impact on nursing home performance due to the change in clinical coding guidelines. Thus, the output suggests that changes to the clinical coding guidelines did not have a substantial effect on provider performance and do not appear to be a threat to the validity of NQF #0686.

Table 3. Distribution of Differences in Facility Performance Scores on NQF #0686 Percent of Residents Who Have/Had a Catheter Inserted and Left in Their Bladder (Long Stay) Across Years, FY 2013 – 2018

					Percentiles									
Difference	*k*	Mean	S.D.	Min	1st	5th	10th	25th	50th	7th	90th	95th	99th	Max
2013 – 2014	14,974	0.002	0.026	-0.30	-0.06	-0.04	-0.02	-0.01	0.00	0.01	0.03	0.04	0.07	0.65
2014 – 2015	14,963	0.001	0.025	-0.44	-0.06	-0.04	-0.02	-0.01	0.00	0.01	0.03	0.04	0.07	0.34
2015 – 2016	14,965	0.004	0.027	-0.56	-0.06	-0.03	-0.02	-0.01	0.00	0.02	0.03	0.04	0.07	0.56
2016 – 2017	14,988	0.005	0.025	-0.53	-0.06	-0.03	-0.02	-0.01	0.00	0.02	0.03	0.04	0.08	0.39
2017 – 2018	14,946	0.003	0.024	-0.35	-0.06	-0.03	-0.02	-0.01	0.00	0.01	0.03	0.04	0.07	0.36

Source: RTI analysis of MDS 3.0 Data, Q4, 2013 through Q3, 2018 (programming reference: il\il61\il61_request_q_31_32_684_fy.log)

1.g. Face validity: The majority of TEP members explicitly affirmed the face validity of NQF #0686. The TEP supported currently specified exclusion criteria of neurogenic bladder and obstructive uropathy, as well as the risk adjustment covariates of pressure ulcers at stages II – IV and frequent bowel incontinence. Some TEP members suggested including additional risk-adjustors but were countered by other members of the TEP, indicating no strong consensus among TEP members. Other TEP members explained the trade-offs of including additional risk adjustors that may not be well documented in the MDS and how that may compromise the validity of the current measure. Overall, TEP members voiced support for the face validity of NQF #0686 as it is currently specified (RTI International, 2019).

RTI International. (2019, June). Technical Expert Panel Summary Report: Maintenance of Nursing Home Quality Measures Prepared under CMS Contract No. HHSM-500-2013-13015I.

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

Performance Measure Score Validity

RTI's analyses indicated that this measure is a valid measurement of catheter use. Facilities' scores on this QM are positively correlated with their scores on NQF #0684 (Percent of Residents with a Urinary Tract Infection (Long Stay)), providing some evidence supporting convergent validity. Although the majority of the variance in NQF #0686 is due to factors other than geographical variation, a small and significant proportion of the variance is explained by the respective states in which nursing facilities are located. This feature of the measure warrants continued monitoring. However, this measure does not vary substantially from quarter to quarter corresponding to changes in seasons; thus, seasonality is not a threat to validity for this measure. Further, the confidence interval analysis for this measure offers evidence of substantial variation among providers and the stability analysis demonstrates the stability of provider performance over time. While some facilities observed shifts of 3 deciles in their performance on NQF #0686 from quarter to quarter, changes in performance of 3 deciles represent a relatively small absolute difference in percentage points between quarters, providing further evidence of intra-facility consistency and stability in provider performance. The 2019 TEP supported the face validity of the measure.

Please see section 2b6 for analysis of the impact of missing data on this measure, which also speaks to validity.

2b2. EXCLUSIONS ANALYSIS

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

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

The denominator exclusion criteria for this quality measure are as follows: 1) The target assessment is an admission assessment, a PPS 5-day assessment or a PPS readmission/return assessment; 2) The target assessment indicates that indwelling catheter status is missing; 3) The target assessment indicates neurogenic bladder or neurogenic bladder status is missing; or 4) The target assessment indicates obstructive uropathy or obstructive uropathy status is missing.

RTI examined the frequency and proportion of residents excluded from this measure for each of the exclusion criteria for this quality measure.

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

A total of 202,813 residents (16.3% of 1,104,988 long-stay residents in Quarter 3, 2018) were excluded from this quality measure based on the measure denominator exclusions as described above.

84,884 residents (6.8% of long-stay residents in Quarter 3, 2018) were excluded because their target assessment indicated that they had an active diagnosis of neurogenic bladder or the item that captures the diagnosis of neurogenic bladder was not completed. The proportion of the long-stay resident population excluded due to a missing or active diagnosis of neurogenic bladder is moderate and the proportion excluded for this reason has remained stable over time. This measure includes only those residents who have had an indwelling catheter in the last seven days. The decision to exclude for neurogenic bladder was comprehensively evaluated through analyses of the MDS 3.0 data set, feedback received from subject matter experts, and based on literature showing that residents with this condition are at risk of catheter use from complications such as overflow incontinence or recurrent bladder infections (Smith, Zheng, Reilly, Kissam, et al, 2012).

35,305 residents (2.8% of long-stay residents in Quarter 3, 2018) were excluded because their target assessment indicated they had an active diagnosis of obstructive uropathy or the item that captures the diagnosis of obstructive uropathy was not completed. The proportion of the long-stay resident population excluded due to a missing active diagnosis of obstructive uropathy is relatively modest. Obstructive uropathy occurs when bladder impairments or abnormalities exist, which may result in the individual not being able to empty their bladder voluntarily or effectively (Smith et al, 2012). Therefore, it was deemed appropriate to exclude for residents with an active diagnosis of obstructive uropathy since the measure captures the percentage of residents who have had an indwelling catheter in the last 7 days.

187 residents (<0.1% of long-stay residents in Quarter 3, 2018) were excluded because data was missing on the item that captured the status of indwelling catheter use.

Tests of the frequency distribution of exclusions across measured entities and impact on performance measure scores are available in section 2b6.

82,237 residents (6.6% of long-stay residents in Quarter 3, 2018) were excluded because their target assessments were either admission, PPS 5-day, or readmission/return assessments (RTI, 2019).

RTI International. (2019). Analysis of Q3 2018 MDS 3.0 data (programming reference: rn27_47\il27_request_686_31_32.log)

Smith, L., Zheng, N. T., Reilly, K., Kissam, S., Rokoske, F., Barch, D., & Manning, J. (2012). Nursing home MDS 3.0 quality measures: Final analytic report.

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

Most denominator exclusions for this measure occur because of a specific active diagnosis or missing data; impact of missing data on this quality measure is presented in detail in section 2b6.

6.8% and 2.8% of long-stay residents were excluded because they had an active diagnosis of neurogenic bladder or obstructive uropathy, respectively, or the items that capture these diagnoses had missing data (RTI, 2019). While this is a fairly moderate proportion of the long-stay resident population, the decision to focus these active diagnoses was previously well vetted using feedback from subject matter experts, the literature showing that residents with these conditions are at risk of catheter use from bladder-related complications, and analyses of the MDS 3.0 data items.

RTI International. (2018). Analysis of Q3 2018 MDS 3.0 data (programming reference: rn27_47\il27_request_q3132_686.log)

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

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

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

☐ **No risk adjustment or stratification**

☒ **Statistical risk model with two risk factors**

☐ **Stratification by** Click here to enter number of categories **risk categories**

☐ **Other,** Click here to enter description

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

This measure is risk-adjusted for bowel incontinence and pressure ulcers at Stage II, III, or IV using a logistic regression. The measure is calculated as follows:

Step 1: Identify the total number of long-stay residents who do not meet the exclusion criteria, with a selected target assessment (OBRA, PPS, or discharge) during the quarter.

Step 2: Calculate the facility-level observed score (steps 2a through 2b below).

Step 2a: Starting with the set of residents identified in Step 1, determine the total number of long-stay residents with a selected target assessment that meet the numerator inclusion criteria.

Step 2b: Calculate the facility observed score by dividing the results of step 2a by the results of step 1.

Step 3: Calculate the national observed score by averaging the scores derived in step 2b across all facilities.

Step 4: Calculate the expected resident score for each resident (steps 4a and 4b below).

Step 4a: Assign covariate values, either '0' for covariate condition not present or '1' for covariate condition present, for the residents included in the denominator for each of the two covariates (bowel

incontinence and presence of pressure ulcers) based on resident's prior assessment and run the logistic regression model.

Specifically, the covariates are calculated as follows:

For the variable identifying frequent bowel incontinence on prior assessment (H0400 = [2, 3]):

1. Covariate = [1] if H0400 = [2, 3];

2. Covariate = [0] if H0400 = [0, 1, 9, -].

For the variable identifying pressure ulcers at stage II, III, or IV in prior assessment:

1. Covariate = [1] if *any* of the following are true:
    a. M0300B1 = [1, 2, 3, 4, 5, 6, 7, 8, 9], *or*
    b. M0300C1 = [1, 2, 3, 4, 5, 6, 7, 8, 9], *or*
    c. M0300D1 = [1, 2, 3, 4, 5, 6, 7, 8, 9].

2. Covariate = [0] if the following is true:
    a. M0300B1 = [0, -, ^] *and*
    b. M0300C1 = [0, -, ^] *and*
    c. M0300D1 = [0, -, ^].

*All covariates are missing if no prior assessment is available.

The logistic regression model is of the form:

[Equation 1] QM triggered (yes=1, no=0) = B0 + B1*COVA + B2*COVB

Where:

B0 is the logistic regression constant (B0 =-4.251405),

B1 is the logistic regression coefficient for the first covariate, bowel incontinence (B1 = 0.4434465),

COVA is the resident-level score for the first covariate (0 or 1),

B2 is the logistic regression coefficient for the second covariate, pressure ulcers at stage II, III, or IV (B2 = 2.277961, and

COVB is the resident-level score for the second covariate (0 or 1)

Step 4b: Calculate the expected resident score for each resident with the following formula:
[Equation 2] Resident-level expected QM score = 1/ [1+e^-x]

Where e is the base of natural logarithms and x is a linear combination of the constant and the logistic regression coefficients times the covariate scores (from Equation [1]). A covariate score will be 1 if the covariate is triggered for that resident, and 0 if the covariate is not triggered.

Step 5: Calculate the facility-level expected QM score by averaging all resident-level expected scores derived in step 4b.

Step 6. Calculate the facility-level adjusted score based on the:

- facility-level observed QM score (step 2b),
- facility-level expected QM score (step 5), and
- national average observed QM score (step 3).

The calculation of the adjusted score uses the following equation:

[Equation 3] Adj = 1/ [1 + e^ -y]

where

Adj is the facility-level adjusted QM score, and

y = (Ln(Obs/(1–Obs) - Ln(Exp/(1–Exp) + Ln(Nat/(1–Nat))

Obs is the facility-level observed QM rate,

Exp is the facility-level expected QM rate,

Nat is the national observed QM rate (Nat = 0.0217076), and
Ln indicates a natural logarithm.
e is the base of natural logarithms

RTI International. (2018). Analysis of Q3, 2018 MDS 3.0 data (programming reference: rn27_47\LJC10_request_q2829_686.log)

Reference: The Centers for Medicare & Medicaid Services (CMS) (April 2017). MDS 3.0 Quality Measures User's Manual. RTI International, Waltham, MA. Accessed at: https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/NursingHomeQualityInits/Downloads/MDS-30-QM-Users-Manual-V11-Final.pdf

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

This is not applicable. This measure is risk-adjusted.

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

This quality measure reports the risk-adjusted percentage of residents who have/had a catheter inserted and left in their bladder among long-stay nursing facility residents. The goal of risk adjustment is to control for differences across facilities in resident characteristics that might be related to the outcome of interest. This allows outcomes to be compared across facilities after differences in resident complexity (i.e., resident characteristics) have been accounted for in the analysis. The risk adjustment model for this measure controls for variation across facilities in the percentage of residents with bowel incontinence and the percentage of residents with pressure ulcers at Stage II, III, or IV. Although literature on patient-level factors influencing catheter use is limited, prior RTI analyses conducted on this measure suggest differences in catheter use between residents with and without bowel incontinence and with and without pressure ulcers at Stage II, III, or IV. Thus, this measure is risk-adjusted to allow percentage of residents who have/had a catheter inserted and left in their bladder to be compared across nursing facilities, which may have differing proportions of residents with bowel incontinence or pressure ulcers at Stage II, III, or IV.

Risk Adjustor Selection – Conceptual Rationale and Statistical Testing

Clinical Risk Factors

Past RTI analysis identified potential clinical risk adjustors, selected based on literature review and clinical relevance. The final model (which includes covariates for bowel incontinence and pressure ulcers at Stage II, III, or IV) was developed after assessing the impact of all identified potential covariates and subsets of the covariates on risk-adjusted scores and on changes in facility scores and ranks. During the recent TEP, some members suggested that hospice care would be a beneficial risk-adjustor for NQF #0686. However, there was not consensus among the rest of the TEP members and some members contested that catheterization increases comfort for hospice residents as there has not been any evidence to support that claims. RTI performed some additional testing to determine the appropriateness of including hospice care as a risk-adjustor.

While there are significant differences in triggering the numerator when comparing residents with and without a prior hospice stay (3.8% and 2.1%, respectively ($p<0.001$)), subsequent analyses found that the performance of providers was not influenced by hospice status. Stability in intercept and covariates, as well of model goodness of fit statistics, over time, supports the validity of the current specification (Smith, Zheng, Reilly, Kissam, et al, 2012).

Social Risk Factors

We conducted a recent literature review related to this quality measure, including topics related to catheter use and urinary continence care in nursing homes, to determine whether other researchers had posited a conceptual basis for why social risk factors might influence the incidence of catheter use in nursing homes, such that the risk factor could not be addressed through nursing home care delivery (NQF, 2017). Some studies have found an empirical association between social risk factors that could be measured by items available in the MDS 3.0 and UTI, a related quality construct, but did not offer a conceptual basis for understanding how the inherent characteristics of the social risk factor (race/ethnicity, age, gender, and Medicaid status) would affect catheter use/non-use. The 2019 TEP did not find a conceptual basis for risk adjustment by any of these social risk factors (RTI International, 2019).

In the event that there is interest in statistical testing on social risk factors with an *empirical* association with the outcome – even in the absence of a *conceptual* reason for the social risk factor—we examined (1) the feasibility of stratifying the measure by race/ethnicity, gender, and Medicaid status, as that would be the most appropriate risk adjustment strategy to avoid masking disparities in care associated with those factors, and (2) the effect of age (equal or greater than 85 years old) in a risk adjustment model in addition to the clinical risk adjustors included in the current specification.

Risk Adjustor Included in the Final Model

The risk adjustors selected for inclusion in the final model, bowel incontinence and pressure ulcers at Stage II, III, or IV, are described in detail in 2b3.1.1.

National Quality Forum (2017, July). Evaluation of the NQF Trial Period for Risk Adjustment for Social Risk Factors. Final Report. https://www.qualityforum.org/Publications/2017/07/Social_Risk_Trial_Final_Report.aspx

RTI International. (2019, June). Technical Expert Panel Summary Report: Maintenance of Nursing Home Quality Measures Prepared under CMS Contract No. HHSM-500-2013-13015I.

Smith, L., Zheng, N. T., Reilly, K., Kissam, S., Rokoske, F., Barch, D., & Manning, J. (2012). Nursing home MDS 3.0 quality measures: Final analytic report.

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

☒ **Published literature**

☒ **Internal data analysis**

☐ **Other (please describe)**

Published literature

We did not develop a conceptual model of how social risk impacts this outcome because there is no rationale presented in the literature for how factors such as race/ethnicity, age, and Medicaid status are influencing the incidence of catheter use in nursing homes. In the case of gender, the association between gender and catheter use is likely reflecting other gender-specific conditions for which facilities should be held accountable when providing care.

The following studies found an *empirical* association between social risk factors that could be measured by items available in the MDS 3.0 and urinary tract infection (a related condition), but did not offer a *conceptual* basis for understanding how the inherent characteristics of the social risk factor (race/ethnicity, age, gender, and Medicaid status) would affect catheter use. One older study (Rogers et al., 2008) discusses a conceptual basis for using gender as a *clinical* risk adjuster. However, as noted above, it is possible that the association between gender and catheter use is reflecting other gender-specific conditions for which facilities should be held accountable when providing care. Similarly, studies looked at the association between age and risk of catheterization and found that older adults are more likely to be catheterized due to a higher likelihood of experiencing adverse health outcomes when being transferred from acute care (i.e., hospital) to SNFs (Burke et al., 2018). Additional studies of relationships between social risk factors and urinary tract infections (as a related condition) include:

- One study identified a positive association between female gender and the rate of urinary tract infections (Gucwa et al. 2016).
- One study identified a link between both age and having Medicaid coverage with urinary tract infections, with older residents having a higher risk of infection and higher Medicaid coverage in a facility negatively associated with urinary tract infection incidence (Castle et al., 2017).
- White race was also identified as a predictor of urinary tract infection in one study (Hefele et al. 2017).

Internal data analysis

We created binary variables for each social risk factor described above as follows:

- Race/ethnicity: defined from item A1000 (Race/Ethnicity) in the MDS. We created non-Hispanic white and non-white categories. A resident is defined as non-Hispanic white if A1000 = F and no other categories apply. A resident is defined as non-white if A1000 was coded as anything other than F.
- Oldest old: defined from Item A0900 (Birth Date) in the MDS. Oldest old is defined as 1 if the resident is age 85 or older and 0 if otherwise. Birth Date is not missing on any assessment in the sample.
- Gender: defined from item A0800 (Gender) in the MDS. Male is defined as 1 and Female as 0. Gender is not missing on any assessment in the sample.
- Medicaid eligibility: defined from Item A0700 (Medicaid Number) in the MDS. Medicaid eligibility is defined as 1 if the resident has a Medicaid number or if a Medicaid number is pending, 0 if Medicaid number = "N", and missing if Medicaid number is missing.

We also used a non-binary version of the race/ethnicity variable, using each of the race/ethnicity categories as defined in item A1000 in the MDS and an additional category for multi-race. Residents were defined as multi-racial if more than one category in item A1000 was selected.

First, we examined the percentage of long-stay residents with each social risk factor identified in the literature as having an empirical association with catheter use, compared to those without that social risk factor, and used Chi-Squared tests to determine whether these differences were statistically significant, as shown in ***Table 4***.

While all differences were statistically significant and absolute percentages are small, the differences across most subpopulations are relatively large. Among residents who are eligible for Medicaid, 2.0% have/had a catheter inserted and left in their bladder and, among those ineligible for Medicaid, 3.0% have/had a catheter inserted ($\chi^2(1) = 472.3$, $p < 0.001$). For residents aged 85 years or older, 1.8% have/had a catheter inserted and left in their bladder, compared to 2.4% of younger residents ($\chi^2(1) = 474.8$, $p < 0.001$). Whereas 3.3% of the male residents have/had a catheter inserted and left in their bladder, 1.6% of the female residents have/had a catheter inserted ($\chi^2(1) = 3{,}200$, $p < 0.001$). Alternatively, the difference across race/ethnicity subpopulations is relatively small. Among residents who are non-Hispanic white, 2.1% have/had a catheter inserted and left in their bladder and, among those who are non-white, 2.3% have/had a catheter inserted ($\chi^2(1) = 33.7$, $p < 0.001$).

Table 4. NQF #0686 Percent of Residents Who Have/Had a Catheter Inserted and Left in Their Bladder (Long Stay) by Social Risk Factors

Resident characteristic (% of all long stay residents)	Frequency of residents who have/had a catheter inserted and left in their bladder	Percentage of residents who have/had a catheter inserted and left in their bladder	Pearson chi^2 P-value
Age (1,033,727)			
≥ 85 (39.93%)	7,383	1.8%	<0.001
< 85 (60.07%)	15,066	2.4%	
Gender (1,033,727)			
Male (34.10%)	11,601	3.3%	<0.001
Female (65.90%)	10,848	1.6%	
Medicaid (947,084)			
Medicaid (88.31%)	16,378	2.0%	<0.001
Non-Medicaid (11.69%)	3,265	3.0%	
Race/ethnicity (1,033,727)			
Non-Hispanic white (73.67%)	16,160	2.1%	<0.001
Non-white (26.33%)	6,289	2.3%	

Source: RTI analysis of Q3, 2018 MDS 3.0 Data (programming reference: LJC67/LJC09_request_686_31_32_rev.log)

Overall, individuals who were younger, male, non-Medicaid eligible, and identified as non-white were more likely than their counterparts to have a catheter inserted and left in their bladder. For males, this result was expected, but other results were inconsistent with empirical findings from the literature.

Given CMS's guidance to avoid having risk adjusters mask disparities in care associated with these factors, and instead using measure stratification by these categories if there is a *conceptual* reason to do so (CMS, 2018), RTI has further examined the implications of stratifying by social risk factors, including race/ethnicity, ,gender, and Medicaid status. Results are shown in ***Table 5***.

When RTI examined race/ethnicity, gender, and Medicaid eligibility as potential stratifying variables for the LS UTI measure, results indicated that, of the facilities with publicly reportable scores (≥20 residents in the denominator) for the current specification, approximately 55.91%, 76.87%, and 93.60% of facilities

would be excluded if the measure were stratified by race/ethnicity, gender, and Medicaid eligibility, respectively. The loss of ability to report the LS Catheter QM would have an effect on its importance and usability in helping consumers (including residents and their caregivers and family) make informed decisions about their nursing home care and in encouraging nursing homes to improve quality in this domain, and thus risk adjustment by stratification is not feasible for this measure.

Table 5 Frequency and percentage of facilities that can report a stratified measure for NQF #0686, Percent of Residents Who Have/Had a Catheter Inserted and Left in Their Bladder (Long Stay)

Social Risk Factor	Number of facilities who can report both levels of the stratified QM	Percentage of facilities who can report both levels of the stratified QM
Race/ethnicity	3,531	23.13%
Sex/gender	6,731	44.09%
Medicaid status	971	6.40%

Source: RTI analysis of Q3, 2018 MDS 3.0 Data (programming reference: LJC70/LJC70_request_686_31_32.log)

Castle, N., et al. (2017). "Resident and Facility Factors Associated With the Incidence of Urinary Tract Infections Identified in the Nursing Home Minimum Data Set." J Appl Gerontol 36(2): 173-194.

Gucwa, A. L., et al. (2016). "Correlations between quality ratings of skilled nursing facilities and multidrug-resistant urinary tract infections." Am J Infect Control 44(11): 1256-1260.

Hefele, J. G., et al. (2017). "Examining Racial and Ethnic Differences in Nursing Home Quality." Jt Comm J Qual Patient Saf 43(11): 554-564

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

The current risk adjustment model includes two covariates: (1) bowel incontinence and (2) pressure ulcers at Stage II, III, or IV. Both of these covariates are significant predictors of catheter use. For the covariate risk adjustment (indirect standardization), the measure is adjusted for bowel incontinence rated as "frequently" (H0400 = [2]) or "always" (H0400 = [3]), and presence of pressure ulcers at Stage II (M0300B1 = [1, 2, 3, 4, 5, 6, 7, 8, or 9]), III (M0300C1 = [1, 2, 3, 4, 5, 6, 7, 8, or 9]), or IV (M0300D1 = [1, 2, 3, 4, 5, 6, 7, 8, or 9]), as coded on the prior MDS assessment.

First, using data from Quarter 3, 2018, we examined the percentage of long-stay residents with bowel incontinence who have/had a catheter inserted and left in their bladder, compared to those who were continent, and the percentage of long-stay residents with pressure ulcers at Stage II, III, or IV who have/had a catheter inserted and left in their bladder, compared to those without pressure ulcers at Stage II, III, or IV. We then used Chi-Squared tests to determine whether these differences were statistically significant, as shown in ***Table 6***. There is a significant difference in the proportion of residents who have/had a catheter inserted and left in their bladder by bowel incontinence: among residents who are bowel incontinent, 2.8% have/had a catheter inserted and left in their bladder, and among residents who are bowel continent, 1.5% have/had a catheter inserted and left in their bladder ($\chi2(1) = 1,900$, $p < 0.001$). There is also a large, significant difference in the proportion of residents who have/had a catheter inserted and left in their bladder by presence of pressure ulcers at Stage II, III, or IV: among residents who have pressure ulcers at Stage II, III, or IV, 15.4% have/had a catheter inserted and left in their bladder, and among residents who do not have pressure ulcers at Stage II, III, or IV, 1.7% have/had a catheter inserted and left in their bladder ($\chi2(1) = 3,100$, $p < 0.001$).

Table 6. NQF #0686 Percent of Residents Who Have/Had a Catheter Inserted and Left in Their Bladder (Long Stay) and Presence of Pressure Ulcers at Stage II, III, or IV, Quarter 3, 2018

Resident characteristic (% of all long stay residents)	Frequency of residents who have/had a catheter inserted and left in their bladder	Percentage of residents who have/had a catheter inserted and left in their bladder	Pearson chi^2 P-value
Bowel incontinence (54.1%)	15,388	2.8%	<0.001
Bowel continence (45.9%)	7,045	1.5%	
Stage II, III, or IV pressure ulcers (3.6%)	5,672	15.4%	<0.001
No Stage II, III, or IV pressure ulcers (96.4%)	16,777	1.7%	

n = 1,011,446

Source: RTI analysis of Q3, 2018 MDS 3.0 Data (programming reference: rn27_47/ ljc09_request_686_31_32.log)

Next, we examined the facility-level distribution of the percentage of residents who are bowel incontinent or who have pressure ulcers at Stage II, III, or IV (presented in ***Table 7***). The facility-level mean of residents who are bowel incontinent was 52.2%, with a standard deviation of 17.8%. The facility-level mean of residents with pressure ulcers at Stage II, III, or IV was 3.5%, with a standard deviation of 3.2%.

Table 7. Distribution of NQF #0686 Percent of Residents Who Have/Had a Catheter Inserted and Left in Their Bladder (Long Stay) and Presence of Stage II, III, or IV Pressure Ulcers Across Facilities, Quarter 3, 2018

Resident characteristic	Facilities (*k*)	Mean % of residents	Std dev.	10th percentile	25th percentile	50th percentile	75th percentile	90th percentile
Bowel incontinence	14,374	52.2%	17.8%	29.0%	40.0%	52.7%	64.7%	74.8%
Stage II, III, or IV pressure ulcers	14,374	3.5%	3.2%	0.0%	1.3%	3.0%	5.0%	7.5%

Source: RTI analysis of Q3, 2018 MDS 3.0 Data (programming reference: rn27_47/ ljc09_request_686_31_32.log)

In the risk adjustment model for the current specification (*n*=1,033,418), the odds ratio for bowel incontinence is 1.56 and is statistically significant at the 0.05 level [95% CI = [1.51, 1.60]]. The odds of having a catheter inserted and left in the bladder are 1.56 times the odds among those who are bowel incontinent, compared to those who are bowel continent. The odds ratio for pressure ulcers at Stage II, III, or IV is 9.66 and is statistically significant at the 0.05 level [95% CI = [9.35, 9.98]]. The odds of having a catheter inserted and left in the bladder are 9.66 times the odds among those with pressure ulcers at Stage II, III, or IV, compared to those without pressure ulcers at Stage II, III, or IV.

The facility-level mean unadjusted percentage of long-stay residents who have/had a catheter inserted and left in their bladder was 2.40% while the risk adjusted facility-level mean was 2.38% (prior to applying public reporting restrictions for facility size). Although the overall facility-level means for the risk-adjusted and unadjusted quality measures are not substantially different, statistical evidence (including the relatively large and statistically significant odds ratios for the covariates, as well as the model fit statistics described in 2b3.6) indicate that this risk adjustment model is valid.

Results of the statistical analyses to examine social risk factors as potential risk adjustors are detailed in 2b3.4b

Source: RTI analysis of Q3, 2018 MDS 3.0 Data (programming reference: rn27_47/LJC10_request_q3132_686.log)

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

For age, the remaining social risk factor with an empirical association with catheter use, RTI examined whether there is variation in the social risk factor among nursing homes (there is; see ***Table 8***), and the potential improvement in the risk model if social risk factors are included. The facility-level mean of residents whose age is greater than or equal to 85 was 41.5%, with a standard deviation of 19.8%.

Table 8. Distribution of Residents Age 85 and Older Across Facilities Reporting NQF #0686 Percent of Residents Who Have/Had a Catheter Inserted and Left in Their Bladder (Long Stay)

Resident characteristics	Facilities (*k*)	Mean % of residents	Std dev.	10th percentile	25th percentile	50th percentile	75th percentile	90th percentile
Age ≥ 85	14,374	41.5%	19.8%	15.5%	27.1%	40.9%	55.6%	68.3%

Source: RTI analysis of Q3, 2018 MDS 3.0 Data (programming reference: rn27_47/LJC09_request_q3132_686.log)

As the analysis in Table 4 above indicates, age of less than 85 years was associated with catheter use and the difference in the proportion of residents who trigger the numerator by age group is relatively large (1.8% vs. 2.4%) and statistically significant. Coupled with the wide range of percentage of oldest old residents, we further considered age of 85 years and older as a potential risk adjustor for this measure. We ran one model risk adjusted for bowel incontinence and pressure ulcers at Stage II, III, or IV (covariates in current risk adjustment model specification), and oldest old.

To ensure accurate comparisons across alternate model specifications, we ran the candidate model and the current specifications using the same sample of residents (*n* = 1,033,418), restricted to those residents with non-missing values for all risk adjustors, including bowel incontinence, pressure ulcers at Stage II, III, or IV, and age of less than 85 years. In the candidate model, the model that additionally adjusts for oldest old, , the odds ratio for bowel incontinence is 1.57 [95% CI = [1.52, 1.61]], the odds ratio for pressure ulcers at Stage II, III, or IV is 9.53 [95% CI = [9.22, 9.84] and the odds ratio for oldest old is 0.76 [0.74, 0.78]. All odds ratios are statistically significant at the 0.05 level.

Similar to the results of the current risk adjustment model, the odds of having a catheter inserted and left in the bladder are 1.57 times the odds among those with bowel incontinence, compared to those with bowel continence; and the odds of having a catheter inserted are 9.53 [95% CI=[9.22, 9.84]] times the odds among those with pressure ulcers at Stage II, III, or IV, compared to those without pressure ulcers at Stage II, III, or IV. Odds of having a catheter inserted are 24% lower among residents aged 85 and older, compared to younger residents.

We then compared the model fit and calibration for this candidate model to those for the current risk adjustment specification. When we additionally risk adjust for oldest old, the c-statistic increases slightly, by 0.01 (from 0.65 to 0.66), compared to the c-statistic for the current risk adjustment model.

Table 9 provides further comparison of the current risk adjustment model and the candidate model specification.

Table 9. Comparison of Model Performance for Alternate Risk Adjustment Specifications, NQF #0686 Percent of Residents Who Have/Had a Catheter Inserted and Left in Their Bladder (Long Stay)

Model Covariates	Hosmer-Lemeshow Chi^2, *P*-value	AIC	BIC	Pseudo R^2	Log Likelihood
Current specification	12.76, <0.001	200,973.9	201,009.5	0.0705	-100,483.9
Candidate specification	182.19, <0.001	200,603.2	200,650.6	0.0722	-100,297.6

n = 1,033,418

AIC: Akaike Information Criterion

BIC: Bayesian Information Criterion

Candidate specification: Risk adjusted for bowel incontinence, pressure ulcers at Stage II, III, or IV, and oldest old

Source: RTI analysis of Q3, 2018 MDS 3.0 Data (programming reference: LJC69/LJC10_request_686_rev.log)

A risk adjustment model that adjusts for the age of residents had acceptable performance, but improvements to model fit and calibration achieved by including this risk factors were very small. Due to the results of our analyses, we do not recommend risk adjusting this measure for social risk factors. There is almost no practical improvement in including this risk adjustor in the model.

In addition, we examined the correlation between quality measure scores, comparing the quality measure scores under the risk adjustment model specification for the candidate model (adjusted for bowel incontinence, pressure ulcers at Stage II, III, or IV, and oldest old) to the current risk adjustment model (adjusted for bowel incontinence and pressure ulcers at Stage II, III, or IV). The quality measure scores calculated following the risk adjustment specifications in the candidate model are highly correlated with the quality measure score under the current risk adjustment model (≥0.998, $p < 0.001$, $k = 14{,}374$). The high degree of correlation between the risk-adjusted QM scores suggests that, for most nursing facilities, the current risk adjustment model and alternate model specifications are not significantly different.

Source: RTI analysis of Q3, 2018 MDS 3.0 Data (programming reference: LJC69/LJC10_request_686_rev.log)

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

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

If stratified, skip to 2b3.9

Our risk adjustment model demonstrates reasonable predictive validity for having a catheter inserted and left in the bladder among long-stay nursing home residents. After running each of the logistic regression models in the restricted sample, we conducted regression diagnostics to assess model performance, examining predictive ability, as well as model fit and calibration. We used the c-statistic to examine the discrimination of the statistical risk model, and we used the Akaike Information Criterion (AIC), Bayesian Information Criterion (BIC), and log-likelihood to examine the statistical risk model calibration. We also provide these values for the current model specification in the unrestricted sample and present the receiving operating characteristic (ROC) curve in 2b3.8.

In addition, for both the restricted and unrestricted samples, we randomly divided the quality measure samples into two groups: a development sample and a validation sample. We estimated the risk adjustment model in the development sample and applied the fitted model to the validation sample. We then compared the model performance, fit, and calibration in these two samples. Results are provided in Tables 9 and 10.

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

The risk adjustment model currently used for this quality measure had acceptable predictive power on the data in Quarter 3, 2018. The multivariate logistic regression model with risk adjustment covariates for bowel incontinence and pressure ulcers at Stage II, III, or IV yielded a c-statistic of 0.65 and a pseudo R^2 of 0.071, values typical for risk adjustment models of nursing facility outcomes based on MDS assessment data (Mukamel, 1997).

In addition, we compared the c-statistic in development and validation samples for the current QM specification. Results are provided in ***Table 10***.

Table 10. Comparison of Model Discrimination Statistics in Development and Validation Samples, NQF #0686 Percent of Residents Who Have/Had a Catheter Inserted and Left in Their Bladder (Long Stay)

	N	Group	C-statistic
Current specifications sub-sample 1	516,709	DEV	0.66
Current specifications, sub-sample 2	516,709	VAL	0.65

Development (DEV)

Validation (VAL)

Source: RTI analysis of Q3, 2018 MDS 3.0 Data (programming reference: LJC/LJC0069_LJC0069_20190517.log)

Mukamel, D. B. (1997). Risk adjusted outcome measures and quality of care in nursing homes. Medical Care, 35, 367–385

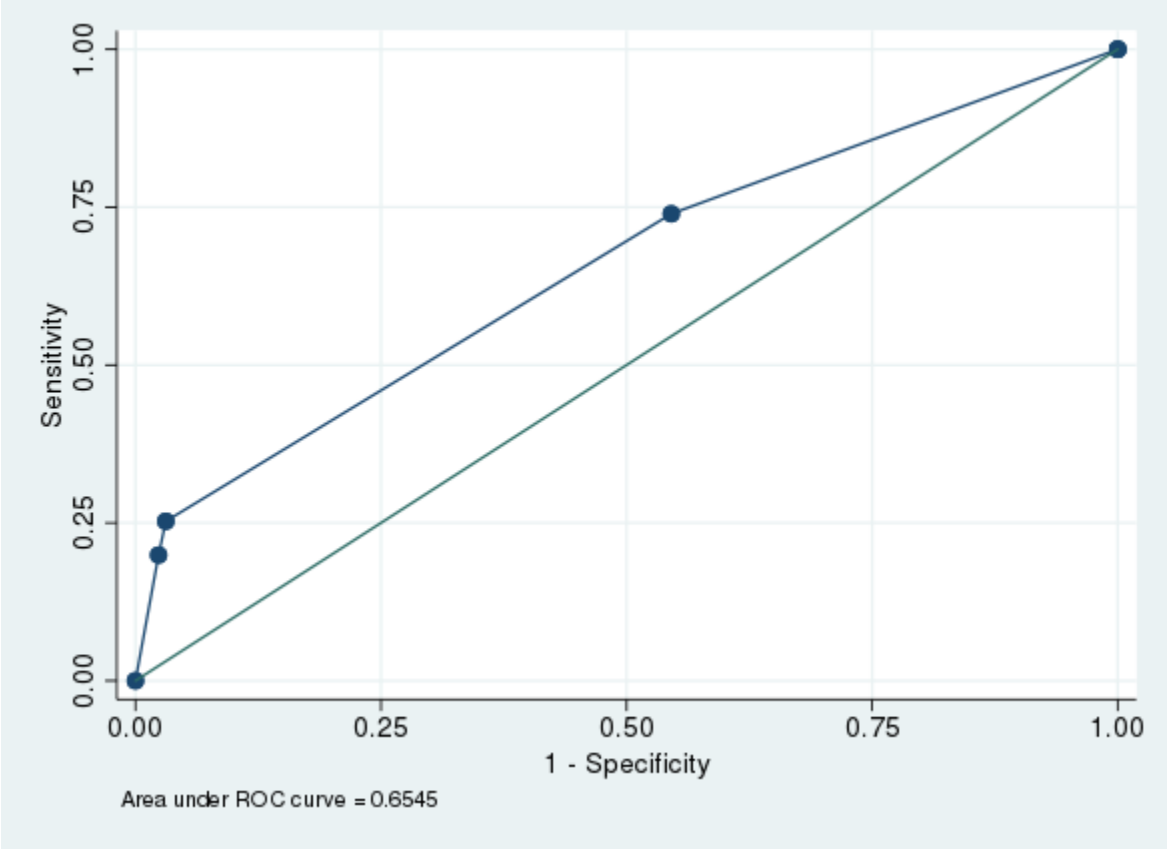
2b3.7. Statistical Risk Model Calibration Statistics (*e.g., Hosmer-Lemeshow statistic*):

The risk adjustment model currently used for this quality measure performed similarly to models with additional risk adjustors when using the restricted sample and unrestricted sample from Quarter 4, 2017. ***Table 9*** above presents the Akaike Information Criterion (AIC), Bayesian Information Criterion (BIC), and log-likelihood for the current risk adjustment model. In general, when comparing models, smaller values for AIC or BIC indicate better model fit; thus, models with smaller values for the information criterion are considered preferable.

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

The receiver operating characteristic (ROC) curve is presented in ***Figure 3*** below. With two binary predictors, the ROC curve is not smooth. Using the current model specification, the area under the curve indicates acceptable model performance. Using the candidate specification (includes oldest old) performs only slightly better than the current specification.

Figure 3. ROC Curve, NQF #0686 Percent of Residents Who Have/Had a Catheter Inserted and Left in Their Bladder (Long Stay), Quarter 3, 2018



Source: RTI analysis of Q3, 2018 MDS 3.0 Data (programming reference: LJC69/LJC10_request_686_rev.log)

2b3.9. Results of Risk Stratification Analysis:

Not applicable. This measure is not stratified.

2b3.10. What is your interpretation of the results in terms of demonstrating adequacy of controlling for differences in patient characteristics (case mix)? (i.*e., what do the results mean and what are the norms for the test conducted*)

In summary, our results demonstrate that the current risk adjustment model with clinical risk factors has reasonable predictive power as measured by the ROC curve. A risk adjustment model that adjusts for the age of residents had acceptable performance, but improvements to model fit and calibration achieved by including this risk factors were very small and offered little practical improvement to the risk adjustment model. Due to the results of our analyses, we recommend maintaining the current risk adjustment model, and do not intend to risk adjust this measure for social risk factors. Risk adjustment by stratification on race/ethnicity, gender, and Medicaid status is infeasible.

2b3.11. Optional Additional Testing for Risk Adjustment (*not required, but would provide additional support of adequacy of risk model, e.g., testing of risk model in another data set; sensitivity analysis for missing data; other methods that were assessed*)

2b4. IDENTIFICATION OF STATISTICALLY SIGNIFICANT & MEANINGFUL DIFFERENCES IN PERFORMANCE

2b4.1. Describe the method for determining if statistically significant and clinically/practically meaningful differences in performance measure scores among the measured entities can be identified (*describe the steps—do not just name a method; what statistical analysis was used? Do not just repeat the information provided related to performance gap in 1b)*

Because the computed scores are not estimates, but include all residents who meet the measure criteria, in terms of discriminating performance, the computed scores can be used to make valid comparisons.

To identify meaningful differences in facility performance on NQF #0686, RTI described the current variability in the facility-level quality measure scores (see 2b4.2). RTI also examined proportions of facilities with scores for this measure that are significantly different from the national facility-level mean, stratified by facility denominator size (see 2b1.3). 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 below average, and scores of 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.

2b4.2. What were the statistical results from testing the ability to identify statistically significant and/or clinically/practically meaningful differences in performance measure scores across measured entities? (e.g., *number and percentage of entities with scores that were statistically significantly different from mean or some benchmark, different from expected; how was meaningful difference defined*)

Table 11 describes the current variability in the quality measure scores of facilities nationally. We find that the mean facility-level score for this quality measure was 2.2% in Quarter 3, 2018 with a median score of 1.7%. The interquartile range for this measure was 3.4%. Among facilities who were eligible to publicly report this measure, 35.2% (*k* = 5,055) had perfect scores of 0%.

Table 11. National Facility-Level Score Distribution, NQF #0686 Percent of Residents Who Have/Had a Catheter Inserted and Left in Their Bladder (Long Stay), Quarter 3, 2018

k	Mean score	Std dev.	10th percentile	25th percentile	50th percentile	75th percentile	90th percentile	% of facilities with "perfect scores"	Interquartile range
14,374	2.2%	2.6%	0.0%	0.0%	1.7%	3.4%	5.5%	35.2%	3.4%

NOTES: *k* = number of facilities that meet minimum requirements for public reporting this quality measure.

Source: RTI analysis of Q3, 2018 MDS 3.0 data (programming reference: rn27_47\ac362_request_q3132_686.log)

Table 2 in section 2b1.3 above shows the proportions of facilities that score statistically significantly higher or lower than the national facility-level mean in Quarter 3, 2018. 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 within the facility's 95% confidence interval.

Overall, just above one-third (38.1%) of facilities were significantly different from the national mean in Quarter 3, 2018, indicating that there are meaningful differences in facility-level scores for this measure. We also stratified the data by the facility denominator size to allow us to examine the relationship between facility size and the reliability of facility scores. The proportions of facilities with scores that are significantly different from the national mean vary as a function of the number of residents included in the

denominator for this measure; in general, the percentage of facilities which have scores that are statistically significantly different from the mean decreases with the number of residents.

Source: RTI analysis of Q3, 2018 MDS 3.0 data (programming reference: rn27_47\ac364_request_q3132_686.log)

2b4.3. What is your interpretation of the results in terms of demonstrating the ability to identify statistically significant and/or clinically/practically meaningful differences in performance across measured entities? (i.*e., what do the results mean in terms of statistical and meaningful differences?*)

These analyses show that the quality measure score varies enough to make meaningful distinctions between high- and low-quality facilities. Moreover, the quality measure scores vary enough from the national mean that there are meaningful differences in facility-level scores for this measure.

2b5. COMPARABILITY OF PERFORMANCE SCORES WHEN MORE THAN ONE SET OF SPECIFICATIONS

If only one set of specifications, this section can be skipped.

Note: *This item is directed to measures that are risk-adjusted (with or without social risk factors)* ***OR*** *to measures with more than one set of specifications/instructions (e.g., one set of specifications for how to identify and compute the measure from medical record abstraction and a different set of specifications for claims or eMeasures). It does not apply to measures that use more than one source of data in one set of specifications/instructions (e.g., claims data to identify the denominator and medical record abstraction for the numerator).* ***Comparability is not required when comparing performance scores with and without social risk factors in the risk adjustment model. However, if comparability is not demonstrated for measures with more than one set of specifications/instructions, the different specifications (e.g., for medical records vs. claims) should be submitted as separate measures.***

2b5.1. Describe the method of testing conducted to compare performance scores for the same entities across the different data sources/specifications (*describe the steps—do not just name a method; what statistical analysis was used*)

Not applicable.

2b5.2. What were the statistical results from testing comparability of performance scores for the same entities when using different data sources/specifications? (*e.g., correlation, rank order*)

Not applicable.

2b5.3. What is your interpretation of the results in terms of the differences in performance measure scores for the same entities across the different data sources/specifications? (i.*e., what do the results mean and what are the norms for the test conducted*)

Not applicable.

2b6. MISSING DATA ANALYSIS AND MINIMIZING BIAS

2b6.1. Describe the method of testing conducted to identify the extent and distribution of missing data (or nonresponse) and demonstrate that performance results are not biased due to systematic missing data (or differences between responders and nonresponders) and how the specified handling of missing data minimizes bias (*describe the steps—do not just name a method; what statistical analysis was used*)

Missing data represent a potential threat to the validity of a quality measure. If patterns indicate that certain types of residents tend to have assessments with missing data in ways that influence the calculation of a quality measure, then that measure may not be capturing outcomes for the intended population. Furthermore, if missing data rates vary systematically across facilities, then the ability to

compare facilities on the measure may be compromised. We examined the rate of missing data at both resident-level and facility-level as well as possible relationships between missing data and the scores for this measure.

RTI analyzed the effects of missing data on this measure in the following ways:

1. We report summary statistics for the facility-level distribution of missing data rates for items used in the calculation of the long-stay catheter measure, both overall and stratified by quality measure score quartile.

2. We analyzed whether missing data on the catheter item varied systematically by several resident-level characteristics which are associated with catheter.

    a. Specifically, we examined if missing data on the catheter item (H0100A), the neurogenic bladder item (I1550), and the obstructive uropathy item (I1650) varied systematically on the following characteristics, which our analyses show are related to catheterization: race/ethnicity, age greater than or equal to 85, gender, Medicaid eligibility status, bowel incontinence (H0400), and pressure ulcers at Stage II, III, or IV (M0300B, M0300C, and M0300D).

2b6.2. What is the overall frequency of missing data, the distribution of missing data across providers, and the results from testing related to missing data? (e.g., results of sensitivity analysis of the effect of various rules for missing data/nonresponse; if no empirical sensitivity analysis, identify the approaches for handling missing data that were considered and pros and cons of each)

Table 12 provides summary statistics for an analysis of the distribution of missing data rates for facilities reporting on this measure. Missing data on items used to calculate this measure are rare, with a mean facility-level missing rate of 0.037% and no missing data at all for at least 90 percent of facilities.

Table 12. Distribution of Facility-Level Missing Rate by Measure Score Quartile, NQF #0686 Percent of Residents Who Have/Had a Catheter Inserted and Left in Their Bladder (Long Stay)

Quality Measure Score Quartile	Facilities (k)	Mean	Std dev.	10th	25th	50th	75th	90th
0–25 (Best)†	5,055	.039%	.57%	0%	0%	0%	0%	0%
26–50	2,132	.013%	.17%	0%	0%	0%	0%	0%
51–75	3,594	.038%	.60%	0%	0%	0%	0%	0%
76–100 (Worst)†	3,593	.045%	.77%	0%	0%	0%	0%	0%
Total	14,374	.037%	.60%	0%	0%	0%	0%	0%

Source: RTI analysis of Q3, 2018 MDS 3.0 data (programming reference: rn27_47\IL29_request_686_31_32.log)

Our analysis showed that the average missing data rate across facilities was 0.037% (shown in the "Total" row of ***Table 12***) on items used to construct this measure. Less than 90% of facilities had any missing data at all on items used to construct this measure.

In addition, RTI examined the relationship between missing data for items used to calculate this measure by quality measure score quartile. ***Table 12*** shows the mean facility-level missing rate for items used to calculate this measure is lowest in the second score quartile, and highest (worst scores) in the highest score quartile, and ranges from 0.013%-0.045%. There is also a significant but weak correlation between missing data and quality measure scores ($r = 0.012$, $p < 0.001$).

This analysis addresses the potential concern that missing data in the items used to construct this QM would lead to under-reporting catheter use, resulting in lower (better) scores on this measure. Our analysis demonstrated that this does not appear to be an issue for this measure.

Table 13 summarizes the results of RTI's analysis of whether missing data varied across selected resident characteristics related to having a catheter inserted and left in the bladder. Specifically, we analyzed whether missing data on H0100A, I1550, or I1650 varied systematically on the following characteristics: race/ethnicity, age greater than or equal to 85, gender, Medicaid eligibility status, bowel incontinence, and pressure ulcers at Stage II, III, or IV, and used Chi-Squared tests to determine whether these differences were statistically significant.

Table 13. Frequency of Missing Data by Select Resident Characteristics Among Long-Stay Residents, NQF #0686 Percent of Residents Who Have/Had a Catheter Inserted and Left in Their Bladder (Long Stay)

Any missing data on H0100A, I1550, and I1650	Percent of Residents (%)	Frequency of Residents (n)	Pearson chi² P-value
Medicaid Eligibility			
Yes	0.03%	290	*.729*
No	0.03%	36	
Age			
Age ≥ 85	0.03%	148	*.693*
Age < 85	0.04%	241	
Gender			
Male	0.04%	151	.173
Female	0.03%	238	
Race/Ethnicity			
Non-Hispanic White	0.04%	291	.663
Non-White/Multi-racial	0.03%	98	
Bowel Incontinence			
Yes	0.03%	173	<0.001
No	0.04%	207	
Pressure Ulcers			
Yes	0.02%	11	0.182
No	0.04%	378	

Source: RTI analysis of Q3, 2018 MDS 3.0 data (programming reference: LJC68\LJC68_request_686.log)

Table 13 shows that differences in missing data by bowel incontinence were statistically significant (p<0.05); however, the magnitude of this difference among individuals with or without the presence of bowel incontinence was relatively small. For residents with bowel incontinence, 0.03% had missing data, compared to 0.04% of residents with bowel continence ($\chi^2(1) = 13.238$, $p < 0.001$). All other resident characteristics included in the analysis did not present significant differences in missing data rates on items used in the calculations of this QM.

2b6.3. What is your interpretation of the results in terms of demonstrating that performance results are not biased due to systematic missing data (or differences between responders and nonresponders) and how the specified handling of missing data minimizes bias**?** (i.*e., what do the results mean in terms of supporting the selected approach for missing data and what are the norms for the test conducted; if no empirical analysis, provide rationale for the selected approach for missing data*)

The mean facility-level missing data rate for this measure was quite low (0.037%), and at least 90% of facilities were not missing data on the catheter item. Rates of missing data on items used to construct this QM are very similar among individuals with selected characteristics related to catheterization (gender, age greater than or equal to 85, race/ethnicity, Medicaid eligibility, bowel incontinence, and presence of pressure ulcers at Stage II, III, or IV). Missing data is only weakly correlated with scores for this QM. Overall, missing data do not present a threat to this measure's validity.

3. Feasibility

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

3a. Byproduct of Care Processes

For clinical measures, the required data elements are routinely generated and used during care delivery (e.g., blood pressure, lab test, diagnosis, medication order).

3a.1. Data Elements Generated as Byproduct of Care Processes.

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

If other:

3b. Electronic Sources

The required data elements are available in electronic health records or other electronic sources. If the required data are not in electronic health records or existing electronic sources, a credible, near-term path to electronic collection is specified.

3b.1. To what extent are the specified data elements available electronically in defined fields (*i.e., data elements that are needed to compute the performance measure score are in defined, computer-readable fields*) Update this field for **maintenance of endorsement**.

ALL data elements are in defined fields in electronic clinical data (e.g., clinical registry, nursing home MDS, home health OASIS)

3b.2. If ALL the data elements needed to compute the performance measure score are not from electronic sources, specify a credible, near-term path to electronic capture, OR provide a rationale for using other than electronic sources. For **maintenance of endorsement**, if this measure is not an eMeasure (eCQM), please describe any efforts to develop an eMeasure (eCQM).

This is not applicable; all data elements used to calculate the measure are in defined fields in electronic clinical data. There are no current efforts to develop this measure as an eMeasure.

3b.3. If this is an eMeasure, provide a summary of the feasibility assessment in an attached file or make available at a measure-specific URL. Please also complete and attach the NQF Feasibility Score Card.

Attachment:

3c. Data Collection Strategy

Demonstration that the data collection strategy (e.g., source, timing, frequency, sampling, patient confidentiality, costs associated with fees/licensing of proprietary measures) can be implemented (e.g., already in operational use, or testing demonstrates that it is ready to put into operational use). For

eMeasures, a feasibility assessment addresses the data elements and measure logic and demonstrates the eMeasure can be implemented or feasibility concerns can be adequately addressed.

3c.1. Required for maintenance of endorsement. Describe difficulties (as a result of testing and/or operational use of the measure) regarding data collection, availability of data, missing data, timing and frequency of data collection, sampling, patient confidentiality, time and cost of data collection, other feasibility/implementation issues.

IF instrument-based, consider implications for both individuals providing data (patients, service recipients, respondents) and those whose performance is being measured.

The general data collection method for the MDS 3.0 is currently operational and mandatory for all Medicare/Medicaid certified nursing homes; no issues are anticipated.

CMS provides coding directions for bowel and bladder items in the MDS 3.0 via the RAI Manual and other mediums, such as this YouTube video explaining the MDS 3.0 coding of Section H.

3c.2. Describe any fees, licensing, or other requirements to use any aspect of the measure as specified *(e.g., value/code set, risk model, programming code, algorithm)*.

This is not applicable.

4. Usability and Use

Extent to which potential audiences (e.g., consumers, purchasers, providers, policy makers) are using or could use performance results for both accountability and performance improvement to achieve the goal of high-quality, efficient healthcare for individuals or populations.

4a. Accountability and Transparency

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

4.1. Current and Planned Use

NQF-endorsed measures are expected to be used in at least one accountability application within 3 years and publicly reported within 6 years of initial endorsement in addition to performance improvement.

Specific Plan for Use	Current Use (for current use provide URL)
	Public Reporting Nursing Home Compare http://www.medicare.gov/nursinghomecompare/search.html Quality Improvement (external benchmarking to organizations) Certification And Survey Provider Enhanced Reports (CASPER) https://qtso.cms.gov/providers/nursing-home-mdsswing-bed-providers Quality Improvement (Internal to the specific organization) Certification And Survey Provider Enhanced Reports (CASPER) https://qtso.cms.gov/providers/nursing-home-mdsswing-bed-providers

4a1.1 For each CURRENT use, checked above (update for maintenance of endorsement), provide:

- Name of program and sponsor
- Purpose
- Geographic area and number and percentage of accountable entities and patients included
- Level of measurement and setting

? Public Reporting:

? Program and sponsor: Nursing Home Compare/CMS
? Purpose: Consumer information on performance
? Geographic area and number and percentages of accountable entities and residents included: All United States Medicare/Medicaid-certified nursing homes with eligible long-stay residents. In Quarter 3, 2018 there were 15,241 eligible facilities containing 1,118,025 residents eligible for inclusion in the measure (before applying sample size restrictions and without excluding residents that do not have a prior and target assessment); 14,374 facilities (94.3%) containing 1,041,783 residents (93.2%) had sufficient sample sizes (20 or more long-stay residents included in the denominator) to report on this measure after applying all measure exclusion criteria and facility sample size restrictions.

? Quality Improvement with Benchmarking (external benchmarking to multiple organizations):
? Program and sponsor: CASPER/CMS
? Purpose: Quality improvement
? Geographic area and number and percentages of accountable entities and residents included: All United States Medicare/Medicaid certified nursing homes with eligible long-stay residents. In Quarter 3, 2018 there were 15,241 eligible facilities containing 1,118,025 residents eligible for inclusion in the measure (before applying sample size restrictions and without excluding residents that do not have a prior and target assessment).

? Quality Improvement (Internal to the specific organization):
? Program and sponsor: CASPER /CMS
? Purpose: Quality improvement
? Geographic area and number and percentages of accountable entities and residents included: All United States Medicare/Medicaid certified nursing homes with eligible long-stay residents. In Quarter 3, 2018 there were 15,241 eligible facilities containing 1,118,025 residents eligible for inclusion in the measure.

SOURCE: RTI analysis of Q3, 2018 MDS 3.0 data (programming reference: ljc67\ljc09_request_686_31-32_rev.log; rn27_47\ac362_request_q3132_686.log)

4a1.2. If not currently publicly reported OR used in at least one other accountability application (e.g., payment program, certification, licensing) what are the reasons? (*e.g., Do policies or actions of the developer/steward or accountable entities restrict access to performance results or impede implementation?*)

This is not applicable; this measure is publicly reported.

4a1.3. If not currently publicly reported OR used in at least one other accountability application, provide a credible plan for implementation within the expected timeframes -- any accountability application within 3 years and publicly reported within 6 years of initial endorsement. (*Credible plan includes the specific program, purpose, intended audience, and timeline for implementing the measure within the specified timeframes. A plan for accountability applications addresses mechanisms for data aggregation and reporting.*)

This is not applicable; this measure is publicly reported.

4a2.1.1. Describe how performance results, data, and assistance with interpretation have been provided to those being measured or other users during development or implementation.

How many and which types of measured entities and/or others were included? If only a sample of measured entities were included, describe the full population and how the sample was selected.

This quality measure (NQF #0686, Percent of Residents Who Have/Had a Catheter Inserted and Left in Their Bladder (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 Nursing Home Compare website (https://www.medicare.gov/NursingHomeCompare/About/nhcinformation.html).

CMS also publishes composite quality ratings on Nursing Home Compare via the Five-Star Rating System.

Further, providers have an opportunity to review their performance prior to public reporting on the Nursing Home 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. These reports also allow providers to view their Five-Star rating. 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

4a2.1.2. Describe the process(es) involved, including when/how often results were provided, what data were provided, what educational/explanatory efforts were made, etc.

The CASPER reports are available to providers on-demand with quality measure data updated monthly. Nursing Home 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 technical users' guides (https://www.cms.gov/Medicare/Provider-Enrollment-and-Certification/CertificationandComplianc/Downloads/usersguide.pdf) on how the quality measures are used in the 5-star rating system, as well as a Help Line, which is accessible by telephone and email, to answer provider questions about the NHQI quality measures and reporting requirements.

4a2.2.1. Summarize the feedback on measure performance and implementation from the measured entities and others described in 4d.1.

Describe how feedback was obtained.

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 on the Nursing Home Compare Help Desk and feedback from the provider community on Open Door Forums (ODFs).

To ensure the continued value and efficacy of the measure, RTI convened a Technical Expert Panel (TEP) to obtain input from providers, residents, and caregivers on the importance, validity, and use of two nursing home quality measures: (1) Percentage of Residents with a Urinary Tract Infection (Long Stay) (NQF #0684); and (2) Percentage of Residents Who Have/Had a Catheter Inserted and Left in Their Bladder (Long Stay) (NQF #0686). The following paragraph outlines how TEP members were recruited and supporting documentation they received to facilitate discussion during the web-based TEP meeting.

On February 12, 2019, RTI posted a Call for TEP Nominations and a TEP Nomination Form on the CMS website to initiate recruitment of TEP members. At the close of the nomination period, RTI finalized the TEP composition by selecting 11 nominees who offered a diverse range of experience, including genitourinary health and care in older adults and nursing homes, consumer perspectives, health care disparities,

performance measurement, quality improvement, and purchaser perspective. Before the TEP meeting, the TEP members received materials to review and complete to prepare for the discussion. Included in these materials was a pre-TEP survey and supplementary materials to assess the TEP members' initial thoughts on the two measures. The pre-TEP survey asked for TEP members' input on focus areas, including the importance, validity, and current use of the two measures. Responses from all TEP members were received before the TEP meeting. De-identified feedback from the TEP members was presented during the TEP meeting on May 23, 2019 and used to inform discussion topics for the TEP meeting.

SOURCE: RTI International. (2019, June). Technical Expert Panel Summary Report: Maintenance of Nursing Home Quality Measures Prepared under CMS Contract No. HHSM-500-2013-13015I.

4a2.2.2. Summarize the feedback obtained from those being measured.

In a pre-TEP survey, TEP members were asked to rate the importance of NQF #0686 on a scale from 1-5 (higher scores are better) based on the following criteria : is an established priority area (National Quality Strategy); addresses a demonstrated high-impact aspect of health care (e.g., affects large numbers); has external evidence of importance, such as consensus standards; and has evidence of disparities for the quality domain. 6 out of 10 TEP members rated this measure as "very important" (rating it a 4 or 5), noting that this measure is important to facilitate quality improvements in nursing homes by bringing continued attention to catheterization rates and working towards timely removal of catheters and reducing catheter use where possible. This measure was also described as tracking a critical health outcome and noted the importance of having this measure for educational opportunities so that nursing homes can take actionable improvements in the care offered to residents.

Other TEP members also noted that they use this QM to track facility performance and this measure serves as an opportunity to educate clinical staff about the impact catheter use has on residents' health outcomes and quality of life. This measure also encourages providers to focus on continence care and reduce catheterization for convenience or without clinical justification. TEP members expressed concerns that removing the measure from Nursing Home Compare would lead to an increase in unnecessary catheterizations as nursing homes would be more likely to focus their attention on other care domains that are publically reported.

RTI also sought input on the measure's validity (i.e., that the measure "produces credible (valid) results about the quality of care when implemented"), including feedback on potential measure modifications and whether risk adjustment for social risk factors would be appropriate, in the context of maintaining or improving the validity of this quality measure, in addition to the current risk adjustment applied to this QM for clinical risk factors via exclusions and a statistical model using certain items as covariates. In particular, RTI asked about gender, age, Medicaid coverage, and race/ethnicity as social risk factors for indwelling catheter placement among low-risk nursing home residents.

TEP members affirmed clear support for this measure's validity and there was broad support for maintaining the current

risk adjustment model to avoid making the risk adjustment model too complex. The TEP members also reached consensus that there was no evidence for including any of the social risk factors in the risk adjustment model and the current exclusions are appropriate. In addition, other TEP members explained that additional risk-adjustment could obscure areas for performance improvement in this quality domain.

Other TEP members inquired about the rates of missing data to determine if that may reduce the validity of the measure. However, missing data on any of the items used to calculate this measure are rare and the average facility missing rate of 0.037% demonstrates that missing data do not present a threat to the measure's validity. Last, one TEP member expressed concerns that the measure "…could be construed as somewhat burdensome" because facilities might not modify an MDS assessment for a long-stay resident when a catheter is removed. Although data element validity warrants continued monitoring and this may be an opportunity for provider training, there was no evidence of this found during the MDS 3.0 pilot testing of these items. Other TEP members also pointed out that this should not be an issue given the way the measure is constructed. For

example, when a long-stay resident's catheter is removed, if the episode is ongoing, this will be captured on the next quarterly assessment and the resident will no longer be triggering the numerator criteria.

SOURCE: RTI International. (2019, June). Technical Expert Panel Summary Report: Maintenance of Nursing Home Quality Measures Prepared under CMS Contract No. HHSM-500-2013-13015I.

4a2.2.3. Summarize the feedback obtained from other users

This is not applicable; additional feedback was not received from other users.

4a2.3. Describe how the feedback described in 4a2.2.1 has been considered when developing or revising the measure specifications or implementation, including whether the measure was modified and why or why not.

Feedback described in 4a2.2.2. demonstrates that TEP members viewed the measure favorably with respect to importance, usability and use, and validity. Based on our synthesis of the literature, empirical testing using MDS 3.0 data, and TEP feedback, we assert that changes to the specifications or use of this measure are not warranted at this time. We will continue to monitor stakeholder feedback and conduct environmental scans to support comprehensive review and evaluation of the measure. We will also continue to monitor catheter-associated guidelines to determine if updates or changes affect the measure's performance from one quarter to the next. CMS will continue to take all feedback into account for future measure refinement.

Improvement

Progress toward achieving the goal of high-quality, efficient healthcare for individuals or populations is demonstrated. If not in use for performance improvement at the time of initial endorsement, then a credible rationale describes how the performance results could be used to further the goal of high-quality, efficient healthcare for individuals or populations.

4b1. Refer to data provided in 1b but do not repeat here. Discuss any progress on improvement (trends in performance results, number and percentage of people receiving high-quality healthcare; Geographic area and number and percentage of accountable entities and patients included.)

If no improvement was demonstrated, what are the reasons? If not in use for performance improvement at the time of initial endorsement, provide a credible rationale that describes how the performance results could be used to further the goal of high-quality, efficient healthcare for individuals or populations.

Progress (trends in performance results, number and percentage of people receiving high-quality healthcare)

? Figure A1 in the appendix presents trend of the national mean and median for this measure across all available quarters (Q1 2011 – Q3 2018). The national facility-level mean and median scores have trended steadily downward since Q1 2011, indicating a general improvement in performance over time. The mean score for this measure was 4.39% in Quarter 1 of 2011 and the median score was 3.74%, demonstrating a positively skewed distribution from many high performing facilities. In Quarter 3 of 2018, the mean and median were 2.24% and 1.66%, respectively.

Geographic area and number and percentages of accountable entities and patients included:

? United States Medicare/Medicaid certified nursing homes with eligible long-stay residents. In Quarter 3, 2018 there were 15,241 eligible facilities containing 1,118,025 residents eligible for inclusion in the measure (before applying sample size restrictions and without excluding residents that do not have a prior and target assessment); 14,374 facilities (94.3%) containing 1,041,783 residents (93.2%) had sufficient sample sizes (20 or more long-stay residents included in the denominator) to report on this measure after applying all measure exclusion criteria and facility sample size restrictions.

4b2. Unintended Consequences

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

4b2.1. Please explain any unexpected findings (positive or negative) during implementation of this measure including unintended impacts on patients.

There were no unexpected findings during the testing process of NQF #0686.

4b2.2. Please explain any unexpected benefits from implementation of this measure.

There were no unexpected benefits during the testing process of NQF #0686.

5. Comparison to Related or Competing Measures

If a measure meets the above criteria and there are endorsed or new related measures (either the same measure focus or the same target population) or competing measures (both the same measure focus and the same target population), the measures are compared to address harmonization and/or selection of the best measure.

5. Relation to Other NQF-endorsed Measures

Are there related measures (conceptually, either same measure focus or target population) or competing measures (conceptually both the same measure focus and same target population)? If yes, list the NQF # and title of all related and/or competing measures.

No

5.1a. List of related or competing measures (selected from NQF-endorsed measures)

5.1b. If related or competing measures are not NQF endorsed please indicate measure title and steward.

5a. Harmonization of Related Measures

The measure specifications are harmonized with related measures;
OR
The differences in specifications are justified

5a.1. If this measure conceptually addresses EITHER the same measure focus OR the same target population as NQF-endorsed measure(s):
Are the measure specifications harmonized to the extent possible?

No

5a.2. If the measure specifications are not completely harmonized, identify the differences, rationale, and impact on interpretability and data collection burden.

N/A

5b. Competing Measures

The measure is superior to competing measures (e.g., is a more valid or efficient way to measure);
OR
Multiple measures are justified.

5b.1. If this measure conceptually addresses both the same measure focus and the same target population as NQF-endorsed measure(s):
Describe why this measure is superior to competing measures (e.g., a more valid or efficient way to measure quality); OR provide a rationale for the additive value of endorsing an additional measure. (Provide analyses when possible.)

N/A – there are no competing measures for NQF #0686.

Appendix

A.1 Supplemental materials may be provided in an appendix. All supplemental materials (such as data collection instrument or methodology reports) should be organized in one file with a table of contents or bookmarks. If material pertains to a specific submission form number, that should be indicated. Requested information should be provided in the submission form and required attachments. There is no guarantee that supplemental materials will be reviewed.

No appendix **Attachment:**

Contact Information

Co.1 Measure Steward (Intellectual Property Owner): Centers for Medicare & Medicaid Services

Co.2 Point of Contact: Carol, Schwartz, Carol.Schwartz@cms.hhs.gov, 410-786-0576-

Co.3 Measure Developer if different from Measure Steward: Acumen LLC

Co.4 Point of Contact: Michael, Collier, mcollier@sphereinstitute.org, 650-558-8882-1268

Additional Information

Ad.1 Workgroup/Expert Panel involved in measure development

Provide a list of sponsoring organizations and workgroup/panel members' names and organizations. Describe the members' role in measure development.

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 (AMDA)

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 over two days in January 2009 to review an environmental scan of the current quality measures and to make recommendations regarding their transition from MDS 2.0 to MDS 3.0 and to assess measure reliability and validity.

In addition, to ensure the continued value and efficacy of the measure, RTI convened a Technical Expert Panel (TEP) on May 23, 2019, to obtain input from providers, residents, and caregivers on the importance, validity, and use of two nursing home quality measures: (1) Percentage of Residents with a Urinary Tract Infection (Long Stay) (NQF #0684); and (2) Percentage of Residents Who Have/Had a Catheter Inserted and Left in Their Bladder (Long Stay) (NQF #0686). The TEP report, including TEP member biographies, is available online at https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/MMS/TEP-Current-Panels.html.

RTI International (2009). Transition of Publicly Reported Nursing Home Measures to MDS 3.0 Draft Technical Expert Panel Report.

RTI International. (2019, June). Technical Expert Panel Summary Report: Maintenance of Nursing Home Quality Measures Prepared under CMS Contract No. HHSM-500-2013-13015I. Available at https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/MMS/TEP-Current-Panels.html.

Measure Developer/Steward Updates and Ongoing Maintenance

Ad.2 Year the measure was first released: 2011

Ad.3 Month and Year of most recent revision: 04, 2016

Ad.4 What is your frequency for review/update of this measure? Endorsement maintenance every 3 years; annual maintenance every year.

Ad.5 When is the next scheduled review/update for this measure? 08, 2020

Ad.6 Copyright statement: This is not applicable.

Ad.7 Disclaimers: This is not applicable.

Ad.8 Additional Information/Comments: This is not applicable. No changes have been made to the measure specifications since the last endorsement.