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 #: 0684

Corresponding Measures:

De.2. Measure Title: Percent of Residents with a Urinary Tract Infection (Long Stay)

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

De.3. Brief Description of Measure: This measure reports the percentage of long-stay residents in a nursing home who have a urinary tract infection in the 30 days prior to the target assessment. 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:

Nursing home residents frequently develop infections (Herzig et al., 2017) and, among these, urinary tract infections (UTIs) are the most common (Herzig et al., 2017; Smith et al., 2018). Symptoms of UTIs include fever, painful or difficult urination, increased frequency and urgency of urination, blood in the urine, low abdominal or flank pain or tenderness, and deterioration in mental status (such as increased confusion). Thus, improving quality of care in genitourinary health domains, including reducing UTIs, may improve quality of life for long-stay nursing home residents in part by decreasing associated signs and symptoms, especially those that cause pain or discomfort.

UTIs can also lead to serious corollary outcomes and complications, such as sepsis, hospitalization, emergency department use, delirium, or death (Castle et al., 2017; Wolff, 2016). For example, some residents who develop UTIs develop blood infections, and 10 percent of these patients die within one week (Gould et al., 2009). UTIs are also a leading cause of avoidable hospitalizations (Nelson & Flynn, 2015; Wolff et al., 2017). Additionally, an increasingly prevalent corollary outcome is antimicrobial resistance in UTIs. Antimicrobial resistance proliferates through inappropriate widespread antibiotic treatment of asymptomatic bacteriuria (often misdiagnosed as UTI), which has no benefits for patients and can cause potential harm such as adverse drug reactions, drug resistance, and subsequent antibiotic-resistant infections (Cooper et al., 2019; Crnich & Drinka, 2014). Therefore, reducing UTIs, as well as improving diagnosis and treatment of UTIs, may improve long-stay nursing home residents' health outcomes.

Gaps in Performance in Nursing Homes:

The UTI quality measure is the only measure in the current measure set that addresses infections and, thereby, is a very important indicator of how facilities prevent and manage infections. UTI rates have been linked to modifiable nursing home factors, such as nurse staffing levels and mix, administering medication on time and documenting nursing care, and education in catheterization protocols and infection prevention measures

(Meddings et al., 2017, Trautner et al., 2017; Nelson & Flynn, 2015). This quality measure should encourage nursing homes to direct resources to these care domains.

Since many UTIs are related to catheters, this quality measure provides an additional incentive for the facility to monitor its catheter use (RTI, 2019b). Nursing homes also vary in their adoption of evidence-based practices for UTI prevention. Practices such as adhering to clinical guidelines, keeping the perineal area clean, ensuring hand hygiene, improving management of urinary incontinence, and implementing hydration regimens have been shown to be effective in preventing UTIs (Montoya et al, 2016; Meddings et al., 2017; Wolff et al., 2017). Thus, this quality measure may also promote wider use of these practices.

Gaps in Performance among Specific Groups of Nursing Home Residents:

There is some evidence in the literature of empirical relationships between patient characteristics and UTI rates:

1. One study identified a positive relationship between female gender and UTI rate (Gucwa et al., 2016).

2. White race has been found to be a predictor of UTIs (Hefele et al., 2017; Castle et al., 2017).

3. Older age is associated with higher rates of UTIs (Castle et al., 2017).

4. Higher rates of Medicaid coverage in a facility are negatively associated with UTI rates among non-catheterized residents and positively associated with UTI rates among catheterized residents (Castle et al., 2017).

Overall, these findings are consistent with RTI's analysis, which found that individuals who were of non-Hispanic white race/ethnicity, older, and female were slightly more likely than their counterparts to have a UTI. In addition, our analyses also demonstrated that individuals who were not eligible for Medicaid were slightly more likely than individuals who were eligible for Medicaid to have a UTI.

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 (MDS) 3.0. In the pre-TEP survey, six out of 10 TEP members rated the UTI 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); has a demonstrated high-impact aspect of healthcare (e.g., affects large numbers); has external evidence of importance, such as consensus standards; and/or has evidence of disparities for the quality domain (RTI, 2019b).

The TEP noted that it was important to maintain public reporting of the UTI quality measure as it reflects a critical health outcome that warrants continued attention, citing significant negative outcomes of UTIs on nursing home residents' function, quality of life, and socialization. The TEP believed this quality measure was important for promoting quality improvement for the following reasons:

1. The UTI quality measure has an important role in promoting accountability and tracking urinary tract infection, allowing facilities to "identify patterns and implement solutions."

2. The presence of this quality measure keeps providers' efforts focused on improving accurate diagnosis and can also lead to improved antibiotic prescribing and antibiotic stewardship.

3. The TEP indicated the UTI quality measure encouraged investment in education and training for clinical and direct care staff.

Castle, N., Engberg, J. B., Wagner, L. M., & Handler, S. (2017). Resident and facility factors associated with the incidence of urinary tract infections identified in the nursing home Minimum Data Set. Journal of Applied Gerontology, 36(2), 173-194.

Cooper, D., McFarland, M., Petrilli, F., & Shells, C. (2019). Reducing inappropriate antibiotics for urinary tract infections in long-term care: A replication study. Journal of Nursing Care Quality, 34(1), 16-21.

Crnich, C. J. & Drinka, P. (2014). Improving the management of urinary tract Infections in nursing homes: It´s time to stop the tail from wagging the dog." Annals of Long Term Care: Clinical Care and Aging, 22(9), 32-36.

Gould CV, Umscheid CA, Agarwal R, Kuntz G, Pegues DA; the Healthcare Infection Control Practices Advisory Committee. Guideline for prevention of catheter-associated urinary tract infections 2009. Atlanta: Centers for Disease Control and Prevention, 2009. Available from

https://www.cdc.gov/infectioncontrol/pdf/guidelines/cauti-guidelines-H.pdf.

Herzig, C., Dick, A., Castle, N., & Stone, P. (2016). Policies and practices to reduce urinary tract infections in nursing homes [Abstract]. American Journal of Infection Control 44, S12.

Meddings, J., Saint, S., Krein, S. L., Gaies, E., Reichert, H., Hickner, A.,…Mody, L. (2017). Systematic review of interventions to reduce urinary tract infection in nursing home residents. Journal of Hospital Medicine, 12(5), 356

Montoya, A., Cassone, M., & Mody, L. (2016). Infections in nursing homes: Epidemiology and prevention programs. Clinics in Geriatric Medicine, 32(3), 585-607.

Nelson, S. T., & Flynn, L. (2015). Relationship between missed care and urinary tract infections in nursing homes. Geriatric Nursing, 36(2), 126-130.

RTI analysis of MDS 3.0 episode files for Quarter 1, 2011–Quarter 3, 2018 (programming reference: KH46\hf15_request_684_31_32.log)

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.

Smith, S. N., Greene, M. T., Mody, L., Banaszak-Holl, J., Petersen, L. D., & Meddings, J. (2018). Evaluation of the association between Nursing Home Survey on Patient Safety culture (NHSOPS) measures and catheter-associated urinary tract infections: Results of a national collaborative. BMJ Quality & Safety, 27(6), 464-473.

Trautner, B. W., Greene, M. T., Krein, S. L., Wald, H. L., Saint, S., Rolle, A. J.,...Mody, L. (2017). Infection prevention and antimicrobial stewardship knowledge for selected infections among nursing home personnel. Infection Control & Hospital Epidemiology, 38(1), 83-88.

Wolff, M. L., et al. (2016). An innovative quality assurance activity to reduce urinary tract infection rates in a green house skilled nursing setting." Annals of Long Term Care, 24(10), 17-20.

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 a urinary tract infection within the last 30 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) and who do not meet the exclusion criteria.

S.8. Denominator Exclusions: If the target assessment is an OBRA Admission Assessment, PPS 5-Day Assessment, or PPS Readmission/Return Assessment, the resident is excluded from the denominator for this quality measure. A resident is also excluded if the target assessment indicates that data is missing for the data element assessing urinary tract infection in the last 30 days.

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:** Nov 10, 2014

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

Evidence Summary

The developers described the clinical importance of UTI as an outcome. They also describe data linking structure and process within nursing homes to this outcome. Finally, they provide a systematic review published in 2017 that describes one or more practices that can be implmemented within nursing homes to reduce the incidence of UTI. Along with reducing the use of urinary catheters, it also includes the following recommended interventions from the review:

- Hand hygiene (R)
- Improve general patient hygiene to reduce infection (R)
- Treatment of atrophic vaginitis as UTI prophylaxis (R)
- Interventions to improve management of urinary incontinence (R)
- Implementation of effective infection control program (R)

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:

Additional studies were described in the narrative demonstrating the importance of UTI and that both structural preditors and process predictors were associated with its incidence in nursing homes.

Question for the Committee:

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

Guidance from the Evidence Algorithm

Health outcome measure (Box 1) -> There is one or more healthcare action that can improve performance on this outcome measure (Box 2) -> PASS

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

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

Maintenance measures – increased emphasis on gap and variation

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

The developer presented the distribution of facility-level scores on this quality measure in Quarter 3, 2018. Overall, 1,096,778 long-stay residents in 14,520 nursing homes are included in the analysis. The national facility-level mean score for this measure in Quarter 3, 2018 was 2.8% and the median score was 1.9%, suggesting a slight positive skew. The interquartile range for this measure was 4.2%. 32.3% of facilities had a perfect score of 0.0%.

Disparities

There were several disparities studies reported by the developer:

1. Differences in performance among specific groups of nursing home residents were small but statistically significant. Among residents who were eligible for Medicaid, 2.46% had a urinary tract infection and, among those ineligible for Medicaid, 3.35% had a urinary tract infection (?2(1) =324.39, p < .001). Among non-Hispanic white residents, 2.88% had a UTI, compared with 1.92% of non-white residents (?2(1) =765.46, p < .001). Among residents aged 85 years or older, 2.76% had a UTI, compared with 2.52% of younger residents (?2(1) = 60.50, p < .001). In addition, whereas 2.20% of male residents had a UTI, 2.84% of female residents had a UTI (?2(1) = 405.96, p < .001). Overall, individuals who were non-Hispanic white, older, female, and not eligible for Medicaid were slightly more likely than their counterparts to have a UTI.

2. At the resident-level, we also compared numerator triggering for this quality measure across more granular racial/ethnic groups. The highest percentage of long-stay residents with a urinary tract infection was found in American Indian or Alaska Native residents (3.0%), followed by White residents (2.9%), Black or African American residents (2.6%), Hispanic or Latino residents (2.0%), Native Hawaiian or Other Pacific Islander residents (1.7%), and Asian residents, who had the lowest percentage of long-stay residents with a UTI (1.6%). Using an ANOVA, differences in the proportion of residents with UTI by racial/ethnic group were found to be statistically significant (p<0.001) (RTI, 2019b).

3. At the facility-level, we compared facility performance on this quality measure in facilities with different proportions of non-Hispanic white and non-white residents. We examined differences in the percentage of long-stay nursing home residents with a UTI across two groups: facilities with proportions of non-Hispanic white residents that were greater than or equal to the median proportion (86.8%) among facilities with sufficient sample size to meet minimum public reporting requirements (>=20 episodes in the denominator), and facilities with a smaller proportion of non-Hispanic white residents (i.e., a larger proportion of non-white residents) than the median proportion. Facilities with a higher proportion of white residents had statistically significantly higher rates of UTI (3.3% compared to 2.4%; p<0.0001) (RTI, 2019c).

4. RTI also compared facility scores on this measure in facilities with different proportions of residents eligible for Medicaid. We examined differences in the percentage of long-stay nursing home residents with a urinary tract infection across two groups: facilities with a large proportion (greater than or equal to 75%) of residents eligible for Medicaid among facilities with sufficient sample size to meet the minimum public reporting threshold, and facilities with fewer than 75% of residents eligible for Medicaid. Facilities with a higher proportion of Medicaid-eligible residents had statistically significantly lower rates of UTI (2.7% compared to 3.7%; $p<0.0001$) (RTI, 2019c).

5. Our testing of social risk factors and their relationships to UTIs indicate that some factors (Medicaid eligibility, non-Hispanic white race/ethnicity, younger age, and male sex/gender were associated with lower UTI rates. Although associations with UTI were generally small, continued monitoring of potential disparities in UTI is critical to ensure continued utility of this measure to providers and consumers of nursing home care.

Questions for the Committee:

- Is there a gap in care that warrants a national performance measure?
- Given the disparities data presented by the developer, does this justify risk adjustment or stratification in reporting?

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:

**Evidence supporting this measure is sound

**This outcome measure is certainly important in concept. Coding data is not ideal to define CAUTIs. There is limited data that structure of NH impacts this outcome, but process certainly does in short term catheter use. This is less clear in long term catheter use

**Applies directly - the developer offered updated information to bloster the connection to outcomes and to process/care chnages.

**There is plenty of evidence of steps to take to prevent UTIs. I am not aware of other studies.

**Previously endorsed measure with evidence to support focus

**Outcomes measure – Evidence shows link between structure (nursing home characteristics and adequate staffing resources) and process (adherence to clinical guidelines and appropriate catheter use) to outcomes of care (lower UTI rate and better corollary outcomes). See Figure 1 submitted.

**study in 2017

**evidence is supportive

**I am concerned about the overtreatment of asymptomatic bacteruria with antibiotics. The CDC has information on this.

**Moderate

1b. Performance Gap

Comments:

** Yes data is presented and a performance gap exists

** Yes there is variability in outcome, The reasons for this are not entirely clearly attributed to preventable quality metrics

** There again is evidence of variation and ongoing opportunity across most sites, with some achieving a performance level very high - but not enough of latter to remove practical opportunity.

** Gaps in performance exist but they seem small. Unclear how this measure is helping to improve quality of care without more data over time (2011-2019). 1/3 facilities have a score of 0 (hard to believe); most facilities had scores better than the mean (5,229 versus 513) indicating it may be close to topping out. I was surprised at the low % of UTIs in this measure.

** yes, performance gap provided. 2.8% in q3 2018

** Gap still exist – Decrease in national mean of UTI from 6.2% in Q2 2013 to 2.8% in Q3 2018; yet 1.9% median score in Q3 2018 suggest slight positive skew toward facilities with UTI reported

** variability

** one panel member raises question about continuation not a lot of change over time

** it should be assumed that patients with an indwelling catheter will develop bacteriuria

** Moderate

Disparities:

** Disparities exist with supporting data

** Yes it was assessed. SES should not impact this outcome within a long term care facility. Gender "disparities" were also assessed

** Evidence about many subpopulation variable opportunity shared suggesting disparities exist and are potential targets for the ongoing meaasure.

** The developer looked at the data by many population groups. There were some disparities in results but they didn't think social risk factors were significant. Higher rates: non-Medicaid, white, >85, female

**Yes

** Disparities shown – significantly higher UTI rate in white, older, female, and Medicaid ineligible population.

** multiple examples

** yes white, non medicaid older females at risk

** n/a

** Yes

Criteria 2: Scientific Acceptability of Measure Properties

2a. Reliability: Specifications and Testing

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

Reliability

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

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

Validity

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

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

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

Evaluators: NQF Scientific Methods Panel Subgroup

Methods Panel Review (Combined)

Methods Panel Evaluation Summary:
Testing data

- The data set used for testing was the Nursing Home Minimum Data Set (MDS) 3.0 v1.15.0
- Two studies were used for the analysis:
    - 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 (Saliba & Buchanan, 2008).

  Included 3,822 residents from community nursing homes and 764 residents from VHA nursing homes

    - RTI facility-level analyses of MDS 3.0 data sample included all facilities with sufficient sample size (n ≥ 20 residents) to publicly report this measure in Quarter 3, 2018 (k = 14,520), unless otherwise noted (RTI International, 2019) Included 1,096, 778 long-stay residents

Ratings for reliability: 5 moderate and 1 low → Measure passes with MODERATE rating

- To test critical data element reliability, the developers examined agreement in coding of the relevant MDS items between 'gold standard' (research) nurses and facility nurses.
    - The Kappa for gold-standard to facility-nurse agreement on the MDS 3.0 and MDS 2.0 item was 0.70. Kappa is a statistical measure of inter-rater agreement for qualitative data, ranging from 0.0 to 1.0. A rating of 0.70 is considered "substantial agreement."

- To test reliability of the measure score, the developers performed:
    - A signal-to-noise analysis: The signal-to-noise ratio for this measure was 0.191 ($p < 0.001$) indicating that 19.1% 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)

    - A split-half reliability analysis The split-half correlation for this measure was positive, but the relationship was moderate ($r = 0.42$, $\rho = 0.37$, $p < .001$), and the ICC was 0.42 ($p < .001$)

Ratings for validity: 1 high, 3 moderate, 1 low, 1 insufficient → Measure passes with a MODERATE rating

- To assess validity of the measure score, the developers examined whether a facility's percentile rank on this measure was correlated with its percentile rank on the related quality measures 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 Bowel or Bladder (Long Stay)).

- As additional support for measure validity, the developers also analyzed:
    - Variation by state
    - Seasonal variation
    - Stability over time (change in percentile ranking in consecutive quarters and average change in performance across years)
    - Face validity: a Technical Expert Panel (TEP) provided feedback on the face validity of NQF #0684. TEP members discussed the current measure specifications, potential risk adjustment factors, and the effectiveness of the measure in capturing quality of care to determine the face validity of the measure as it is currently specified.

Methods Panel Concerns:

- One reviewer found the specifications to be unclear with regard to the measurement timeframe.
- Panel members suggested that testing results showed strong reliability at the data element level but only moderate reliability at the measure score level.
- Some concerns were raised about the adequacy of the developers' rationale for not risk adjusting the measure and the lack of any testing of risk adjustment models.

Summarized Information Provided by Patient Safety Project Team:

Reliability

- Critical Data Element Reliability
    - In their testing of the MDS 3.0, RAND calculated the UTI rate using the MDS 3.0 and the MDS 2.0, both at the individual resident-level and at the facility-level (Saliba & Buchanan, 2008). At the resident-level, the UTI rate using the MDS 2.0 was 10.0% and using the MDS 3.0 was 7.5%. At the facility-level, the MDS 2.0 rate of UTIs was 10.2% and the MDS 3.0 rate was 7.3%. Correlation between the MDS 2.0 and MDS 3.0 measures was strong at both the resident- (œÅ = 0.71) and facility-level (œÅ = 0.80). The Kappa for gold-standard to facility-nurse agreement on the MDS 3.0 and MDS 2.0 item was 0.70. Kappa is a statistical measure of inter-rater agreement for qualitative data, ranging from 0.0 to 1.0. A rating of 0.70 is considered "substantial agreement." These results are indicative of data element reliability.
- Performance Measure Score Reliability
    - Signal-to-noise: The signal-to-noise ratio for this measure was 0.191 ($p < 0.001$) indicating that 19.1% 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 moderate ($r = 0.42$, $\rho = 0.37$, $p < .001$), and the ICC was 0.42 ($p < .001$) (RTI International, 2019b). Although approximately one-third of all facilities have values of 0% for this quality measure, this analysis provides moderate evidence of internal reliability because the variation in scores is sufficient: as shown in Table 8 in Section 2b4.2 below, the 50th percentile score is 1.9% and the 90th percentile score is 7.2%.

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 #0684 (Percent of Residents with a Urinary Tract Infection (Long Stay)) and #0686 (Percent of Residents Who Have/Had a Catheter Inserted and Left in Their Bladder (Long Stay)) and found a positive but weak ($\rho = 0.11$, 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 #0684 (Percent of Residents with a Urinary Tract Infection (Long Stay)) and NQF #0685 (Percent of Low Risk Residents Who Lose Control of their Bowels or Bladder (Long Stay)) and found a positive ($\rho=0.03$) and statistically significant ($p<0.001$) 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 #0684. The proportion of variance in this measure explained by the state in which facilities are located is 1.4% ($p < 0.001$). The interquartile range of state-level scores is 4.2% (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. 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, this also indicates that facilities may have improved practices related to genitourinary care, including prevention of urinary tract infections, during this period.
- Stability analysis: Figure 2 illustrates the changes in facility rank by quality measure score from Quarter 2, 2018, to Quarter 3, 2018. Most (48.5%) facilities are in the same decile in both quarters. Shifts of more than 3 deciles were less common, occurring for approximately 28.4% 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 many more facilities with quality measure scores that were statistically significantly ($p \leq .05$) lower than the national mean of 2.83% than those with scores that were statistically significantly higher than the national mean (5,229 versus 513), 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 decreases with the number of residents, except among the largest facilities (9th and 10th) deciles. Increases in the facility-level sample size lead to reductions in the standard error of facility-level scores, but larger facilities might have greater stabilities due to their larger sample size, which is less affected by a single infection. Changes in the reliability of this measure for the larger facilities may be accounted for by the greater statistical reliability that accompanies increased sample size as well as the increased stability.

- Overall, 39.5% 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 providing evidence of validity for NQF #0684.
- Average change in performance across years: Table 3 presents the changes in provider performance scores from year to year, fiscal year (FY) 2014 – FY 2018. On average, provider scores changed by less than 0.01 percentage points on NQF #0684. Few facilities experienced a change in performance by 0.05 percentage points or greater and over 90% of provider scores changed by 0.07 percentage points or fewer 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 0.007 percentage points, which is similar to previous mean facility score changes between other years when there was no change in clinical coding guidelines (0.006 – 0.009). Based on these findings, we include that there was no substantial change in provider score differences between years, including years when clinical coding guidelines did change (considering scores between FY 2017 and FY 2018, as the coding guideline changed at the beginning of FY 2018). 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 #0684.
- Face validity: The majority of TEP members explicitly affirmed the face validity of NQF #0684. The TEP supported continued public reporting of the measure, as it allows providers to track their performance not only in correctly diagnosing UTIs, but also in antibiotic stewardship, which is closely linked to UTI management. Most TEP members agreed that the measure facilitated a declining trend in UTI rates over time, and reflected quality of care in nursing homes. Some TEP members suggested looking at the relationship between UTI and function, hospice care, and dementia to see if they might be appropriate risk adjustors (see Section 2b3.3a for analysis of candidate factors); however, TEP members voiced support for the face validity of NQF #0684 as it is currently specified (RTI International, 2019).

- <u>Performance Measure Score Validity</u>
  - RTI's analyses indicated that this measure is a valid measurement of urinary tract infections. The testing results indicated high validity according to analysis of seasonal variation. Facility-level measure scores do not vary substantially from quarter to quarter corresponding to changes in seasons; thus, seasonality is not a threat to validity for this measure. The testing results also indicate high validity according to analysis of change in measure performance over years, confidence interval analysis, and variation by state (with a low proportion of variance explained by state). The measure also showed moderate validity according to correlations with related quality measures; i.e., facilities' scores on this QM are positively correlated with their scores on #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)), providing some evidence supporting convergent validity. The testing did suggest some instability in facility decile ranking over time, potentially attributable to lower prevalence and narrow distribution. However, change in measure performance over years was small, even after the revision of item coding guidelines, indicating that changes in facility scores are not due to change in coding practice. The 2019 TEP supported the face validity of the measure.

Questions for the Committee regarding reliability:

- The Scientific Methods Panel (SMP) passed the measure on reliability. Does the Committee want to revote or accept the SMP's rating?
- Do you have any concerns that the measure can be consistently implemented (i.e., are measure specifications adequate)?

Questions for the Committee regarding validity:

- The Scientific Methods Panel (SMP) passed the measure on validity. Does the Committee want to revote or accept the SMP's rating?
- 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:
**Appropriate level of reliability and positively reviewed by scientific methods panel.
**The reliability of this metric is a bit troubling in that coding data is being used, which is not a good proxy for clinical UTI or CAUTI. However, provider identified CAUTI, is by itself an important metric related to antibiotic usage. The lack of assessment of risk adjustment (other than new data presented on mobility risk factors) is also slightly concerning. While catheter utilization is more likely to impacted by patient conditions, prolonged use of a catheter is a risk factor for "CAUTI" and there are differences in duration of catheterization that are independent of quality. In the acute care setting the most important risk factor for CAUTI is duration of catheterization, and it is likely that it a non-linear risk, and acute care facilities can impact usage more readily than I would anticipate a LTC facility can in someone with a neurogenic bladder. I would also worry about non-random variation in coding of UTI
**Solid but not spectacular reliability with some concerns about classification of infection.
**There were some concerns regarding how the patients are defined as "long-stay" that I believe needs to be addressed. Unclear to one of the science panelists who raised issues re 101 days at any nursing home? what about hospital stays within the 101 days and I wonder if that would affect the outcomes. It would help to further define the exclusions as those are defined by types of assessments.
**no concerns
**Moderate – Minimum Data Set (MDS) reliability – 0.70 Kappa score considered substantial agreement; Signal-to-noise ratio – 19.1% ($p<0.001$) considered reliable in separating inter-facility characteristics; Split half reliability – correlation positive but relationship moderate ($r=0.42$).
**kappa score is adequate
**moderate one panel member raises concern about risk adjustment and need to continue
**My concern is over testing and over treatment for asymptomatic patients in nursing homes. Also, this does not address the more fundamental issue of overuse of catheters in nursing homes.
**No concerns

2a2. Reliability – Testing

Comments:
**Modest gaps in reliability are outweighed by importance to track UTIs in this population
**please see above detailed response
**Modest reliability as noted above. Score level testing done but new evidence on use is less robust.
**I think it is moderate - would like to see more recent data and comparisons since 2011; this measure has been around a long time and as with all maintenance measures, I wonder why we don't demand more data from the developer, esp measures like these that are publicly reported.
**no
**Any concerns with moderate reliability at the measure score level?
**no but no risk adjustment
**see above for size of problem, little change over time and need for risk adjustment
**see above (in reliability specification comment)

**no

2b1. Validity – Testing

Comments:
**No concerns
**Yes. Coding for UTI is unlikely to "validly" identify clinical UTI.
**Performance data have aagin modest evidence of validity. Face validity seems high, but actual data are less robust.
**no
**no
**Performance measure score validity – high validity shown in no seasonal variation found and small change in performance over time; moderate validity according to correlations with related quality measures
**no
**no
**see above
**no

2b4-7. Threats to Validity

Comments:
**Comfortable with the SMP recommendations (passing on reliability and validity)
**It does not seem like there is missing data. For the stability analysis I would favor using a longer time period. Quarter to quarter variation is more variable (randomly) than yearly or q6 month evaluation within facility to assess stability. There is limited data presented on how this measure relates to meaningful differences in quality.
**No concerns
**I don't think the data shows significant meaningful differences in quality. Not sure where the benefit of the measure is.
**no
**About 30% of facilities reported 0 UTI in Q3 2018
**no
**nothing to add
**should patients be subdivided further (catheter/no catheter/continent/incontinent). Also the measure does not address the way in which specimens are obtained.
**no

2b2-3. Other Threats to Validity
2b2. Exclusions
2b3. Risk Adjustment

Comments:
**No concerns
**I don't think social risk factors are important for this measure. I do think there needs to be more robust risk assessment or better adjustment of at-risk time to be able to compare fairly across LTC institutons
**The risk adjustment seems underexplained/assessed. It seems possible that differences in populations across sites alone could alter performance outside of specific care actions; only question is "how much" and does it undermine measure.
**As stated above, it would help to have a clarification on exclusions since I'm not familiar with all of the Nursing home assessments and my guess is others are not either. I think the developer addressed the issues of risk adjustment (not needed and I agree) and social risk factors (some differences but not strong and certainly not in the direction that those concerned with social risk factors generally worry about.
**appears appropriate

**No risk adjustment
**would be better to have some risk adjustment but not critical
**why not risk adjusted?
**It can be technically difficult to obtain adequate urine samples in this patient population due to dementia, contractures, skin issues, etc.
**Yes included

Scientific Acceptability: Preliminary Analysis Form (Methods Panel Combined)

Measure Number: 0684

Measure Title: Percent of Residents with a Urinary Tract Infection (Long Stay)

Panel Member #5: NOTE: Several items reported in this submission were completed by colleagues of mine at University of Colorado. However, I was not part of the analytic team that conducted these analyses. I believe I can be objective in my review of the materials. However, if NQF staff believe that I should abstain from reviewing this measure, I will accept their determination.

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**

Panel Member #1: Nursing Home Minimum Data Set (MDS) 3.1 v1.15.0

Panel Member #3: Nursing Home Minimum Data Set (MDS) 3.1 v1.15.0

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**

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

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

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

   Panel Member #1: The requirement for 101 days to qualify for the denominator is unclear how these days are identified, e.g.:

   1) It is silent as to when the clock begins to count days that qualify towards the 101 days. Thus, since it's silent it then appears these days could occur at any point in the past for a given case. Such means to track nursing home stays would be challenging & it would need to be discussed how this occurs, but there's no discussion of the methods employed.
   2) It appears days count in any nursing home qualifies to be counted in this number of days. Thus, it's unclear how stays are identified in any nursing home & who is responsible for identifying these stays. Example: If CMS, it does not appear CMS would be aware of all nursing home stays for some would be private pay or private insurance.

3) In S5 (in the MIF) it does note the days are cumulative & that "Residents who return to the nursing home following a hospital discharge will not have their cumulative days in facility reset to zero. " However it's unstatated the look back period to count number of days.

Panel Member #2: No concerns**.**

Panel Member #3: Although missing data is excluded by definition the potential for confounding by missing data is theoretically possible. However, documentation that the percentage of missing data is minimal (0.04%) and has minimal impact on outcome is convincing.

Panel Member #5: I have no concerns about the measaure specifications. No changes in the measure specifications have been made since the measure last underwent maintenance in 2014. The methodology is clearly stated in line with methodologies utilized in other measures applied to long-stay Medicare beneficiaries. Precise specs for denominator exclusions are provided.

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**

   "Critical Data Element Reliability

   …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). … Cohen's kappas, which 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" [p8]

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**

   Panel Member #1: NA – score level reliability test conducted

   "Performance Measure Score Reliability

   2.a. A signal-to-noise analysis was performed to determine what proportion of total variance in the measure is attributable to differences among providers [p8]

   2.b. 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." [p9]

   Panel Member #2: Not applicable.

   Panel Member #5: (NA = X)

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

   Submission document: Testing attachment, section 2a2.2

   Panel Member #1: <u>Re data level testing:</u> A couple of concerns:

   [1] In 2a2.2 it cites a test conducted in 2008 with "MDS 3". However, it's 11 yrs later & some question as to how the current "MDS 3 v.15" is similar or different from the "MDS 3" from 2008 (where the measure stewart neglects to note the version #).

[2] In the testing, the short & long term cases were mixed. However, this measure is for long term cases (only). Would have been preferable for the study to match the measure in this regard.

[3] Unclear why MDS 2 test results were viewed against MDS 3 test results.

Re measure score testing: Testing seems adequate

Panel Member #2: The methods to test the measure's reliability are appropriate.

Panel Member #3: Very thorough: Data element compared facility reporting with those of "gold standard" trained nurse abstractors with correlations and Kappas calculated.

For performance level testing, outside institution (University of Colorado) examined percentage of facilities that had a change in ranking from one quafter to the next for at least three deiles, 2003 Q3 through 2006 Q3.

Similar analysis RTI examined extent relative facility randk changed quarterly

Panel Member #4: The methods used for reliability testing seem generally appropriate at both levels of analysis. Reasonable choices were made to establish reliability of the key numerator data element of presence of urinary tract infection, and reasonable choices were made to assess reliability of scores for facilities based on rates of UTI among residents.

Panel Member #5: Adequate data element reliability methodology

Panel Member #6: The data set used for reliability testing was the Nursing Home Minimum Data Set (MDS) 3.0 v1.15.0. In the original measure, testing was performed by RAND. For this submission, the testing was performed by RTI, specifically the datasets from Q2 and Q3, 2018. The measure score is the proportion of patients who get a UTI in the eligible population, clearly specified both in the original document and this application.

Data elements comprising the MDS 3.0 were tested in the original submission for inter-rater reliability in 2008. The kappa statistic was used to asses inter-rater reliability between gold-standard nurses and trained facilty nurses. Although the referenced Saliba article mentions high reliability, I could not locate Appendix A for the exact value. For this submission, RAND performed the kappa statistic between MDS 3.0 and MDS 2.0 and found a 0.70 kappa for the UTI data element, considered to be "substantial agreement".

For the measure score, the signal-to-noise ratio was 0.191. Split-half testing was moderate and the ICC was 0.42, which was said to be acceptable but less than the acceptable limit for strong reliability. One-third of facilities had values of 0% for the measure score, but variation was noted, with a 50th percentile score of 1.9% and a 90th percentile score of 7.2%.

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

Submission document: Testing attachment, section 2a2.3

Panel Member #1: Re data level testing: Results are adequate.

Re measure score testing: Signal to noise results are poor & indicate a high level of noise in the measure.

Meaning the measure poorly distinguishes a quality signal. The results from the split half correlation tests were generally weak to modest.

"Critical Data Element Reliability

Correlation between the MDS 2.0 and MDS 3.0 measures was strong at both the resident- (ρ = 0.71) and facility-level (ρ = 0.80). The Kappa for gold-standard to facility-nurse agreement on the MDS 3.0 and MDS 2.0 item was 0.70. Kappa is a statistical measure of inter-rater agreement for qualitative data, ranging from 0.0 to 1.0. A rating of 0.70 is considered "substantial agreement." These results are indicative of data element reliability." [p9]

"Performance Measure Score Reliability

The signal-to-noise ratio for this measure was 0.191 ($p < 0.001$) indicating that 19.1% of the variance in scores for this measure in Quarter 3, 2018 was explained by inter-facility characteristics

split-half correlation for this measure was positive, but the relationship was moderate (r = 0.42, ρ = 0.37, p < .001), and the ICC was 0.42 (p < .001) (RTI International, 2019b). Although approximately one-third of all facilities have values of 0% for this quality measure, this analysis provides moderate evidence of internal reliability because the variation in scores is sufficient: as shown in **Table 8** in **Section 2b4.2** below, the 50th percentile score is 1.9% and the 90th percentile score is 7.2%." [p9]

Panel Member #2: The results are interpreted correctly.

Panel Member #3: For data element testing Kappa was 0.70, correlation was 0.80

For Score testing signal to noise ratio was 0.191 (p<0.001)—19.1% of variance was explained by inter-facility differences—not great but probably acceptable. Split-half reliability was moderate r=0.42, p<0.001) and the ICC was 0.42

Panel Member #4: Results generally showed good reliability at the data element level. Results at the measure score level were modest – probably just into the acceptable range, but only for certain analytic purposes. Facilities are generally not assigned to deciles in a reliable way, so any use that involves decisions and actions based on decile-level grouping or anything more fine-grained than that would not be reliable. It does seem that facilities can be identified as significantly above or below average with acceptable reliability, so any endorsement of the measure should point out that there are only certain things that can be done with the measure that would yield reliable results.

Panel Member #5: Adequate data element reliability testing results

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)

   Panel Member #1: See response to Q6 above.

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

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

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

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

    ☐ **Insufficient** (NOTE: Should rate <u>INSUFFICIENT</u> 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.**

Panel Member #1: Specifications are not precise: See response to Q2 above for specifics.

Panel Member #2: Based on the results of the performance score.

Panel Member #3: Testing was thorough and appropriate but moderate results suggest that overall reliability of the test is moderate.

Panel Member #4: Results generally showed good reliability at the data element level. Results at the measure score level were modest – probably just into the acceptable range, but only for certain analytic purposes. Only 19% of the observed variance among facilities was due to "signal", and there will be some concern that at least part of the "signal" is due to inadequate risk adjustment. Facilities are generally not assigned to deciles in a reliable way, so any use that involves decisions and actions based on decile-level grouping or anything more fine-grained than that would not be reliable. It does seem that facilities can be identified as significantly above or below average with modest reliability, so any endorsement of the measure should point out that there are only certain things that can be done with the measure that would yield reliable results.

Panel Member #5: Adequate data element reliability; national data used

Panel Member #6: As acknowledged by the testing organization, the data element had strong reliability by kappa score, and the measure score had moderate reliability by split-half and intraclass correlation coefficients. I have no other concerns.

VALIDITY: ASSESSMENT OF THREATS TO VALIDITY

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

Submission document: Testing attachment, section 2b2.

Panel Member #1: Some concern of the exclusion that states "missing data in the response to urinary tract infection item in the target assessment". Rationale: Exclusion has the unintended consequence of gaming the measure. However, this represents only 0.04% of denominator cases. [p17]

Panel Member #2: No concerns.

Panel Member #3: None

Panel Member #5: The exclusions seem appropriate.

Panel Member #6: The only exclusions were those for whom data was not available or the target assessment did not meet the inclusion criteria. These are explicitly stated and account for 1.9% of the residents with almost all of these under the latter criteria. No correlation was demonstrated between missing data and measure score performance.

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

Submission document: Testing attachment, section 2b4.

Panel Member #1: No concerns

Panel Member #2: No concerns.

Panel Member #3: See comments below regarding risk adjustment

Panel Member #4: The measure developers do not provide any scientific basis for determining what size difference between or among facilities is "meaningful". The measure has some ability to distinguish among facilties at the top and bottom of the score distribution, as noted above. These differences presumably are "meaningful". The measure cannot distinguish among facilities in the middle of the

distribution with acceptable reliability, but one could argue that these differences are not "meaningful" anyway.

Panel Member #5: The original analyses cited are nearly 15 years old. Even the results presented in Table 1 are more than 5 years old. The seasonal variation results are interesting, but trend results (2 years from more than 5 years ago) seem inadequate given the long period of time that this measure has been in use.

Panel Member #6: The documentation provides data to demonstrate meaningful differences in performance, which have persisted when adjusted for multiple factors. Validity was assessed by correlation with other relevant quality measures, state-to-state, seasonality, stability, and by TEP input as to face validity.

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.

Panel Member #1: No concerns as this is not applicable

Panel Member #2: Not applicable.

Panel Member #3: Not applicable.

Panel Member #5: Submitter labeled this "NA"

Panel Member #6: I have no concerns. Multiple methodologies were applied and all supported the validity of the measure score.

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

Submission document: Testing attachment, section 2b6.

Panel Member #1: No concerns as the missing data of the field capturing numerator is 0.04% of the denominator cases. [p28]

Panel Member #2: Weak correlation between missing data and QM score, however, I don't think it's a significant issue for this measure.

Panel Member #3: As noted in #2 above, potential concerns regarding missing data have been well addressed.

Panel Member #5: Submitter labeled this "NA"

Panel Member #6: No concerns. Missing data information is provided as well as performance across those with missing data without discernible differences.

16. **Risk Adjustment 2b3.**

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

Panel Member #1: Some concern in the examination of potential risk variables, it states for each that "the mean proportion does not increase monotonically across QM score deciles". In and of itself, this is not a rationale to discount the variable from risk adjustment. The measure steward should have tested whether these risk factors were significant in a risk model. (The verbiage states the "results from testing…", but fails to provide specifics other than what I cited above.). All said, the range of cases experiencing a UTI across the continuum for a given risk factor was marginal for most potential risk factors. Thus, it seems most/all would not have been significant in testing in a risk model. [p18]

Panel Member #3: There is extensive discussion of this topic. Although there is a claim that stratification by individual factors is not feasible, the demonstration that there is statistical association not only with age but with other potential risk factors suggests that a model which appropriately adjusted for these factors might

improve the performance of the measure. Indeed, the absence of this adjustment may help to account for why only 19% of the variance in test results can be accounted for by variance among institutions. Even though this rationale has been accepted in the past, it is unclear why a demonstration of the impact of risk adjustment on measure performance has not been peformed.

Panel Member #5: (However, in my opinion, the rationale is weak, and even includes an article that identifies catheter use as a variable in UTI onset.)

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

Panel Member #2: Not applicable.

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

Panel Member #5: (Unknown due to lack of testing)

16d.**Risk adjustment summary:**

Panel Member #3: Not applicable—risk adjustment not performed.

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

Panel Member #1: NA-not risk adjusted

Panel Member #2: Not applicable.

Panel Member #5: (NA =X)

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

Panel Member #1: NA-not risk adjusted

Panel Member #2: Not applicable.

Panel Member #5: (NA =X)

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

Panel Member #1: NA-not risk adjusted

Panel Member #2: Not applicable.

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

Panel Member #1: NA-not risk adjusted

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

Panel Member #1: NA-not risk adjusted

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

Panel Member #1: NA-not risk adjusted

Panel Member #3: As noted above, even though there extensive discussion and explanation of the rationale for not performing risk adjustment, fact remains that available variables appear to be correlated with the measure, and appropriate adjustment might improve the performance of the measure.

Panel Member #4: This is the most disappointing feature of this measure in my opinion. The developers show that several factors have a significant association with the occurrence of UTI, and it would seem logical to attempt to use those factors in some form of adjustment model. They only actually test the variable (oldest age) with the weakest univariate association with UTI, and find (not surprisingly) that its use in an adjustment model does not make much difference and the model fit isn't very good. They seem to feel that stratification is some form of adjustment, based on a CMS guideance document, and point out, fairly enough, that the sample sizes in most facilities are too small to support stratification on the basis of several potential risk variables. The absence of any risk adjustment leaves open the possibility that a

significant part of the "signal" observed in the reliability analyses is unadjusted case mix, and therefore the reliability statistics are over-stated. Those statistics are not strong as it is, leaving significant questions about whether the unadjusted measure can really distinguish among facilities except at the most extreme ends of the score distribution.

Panel Member #5: This measure should be risk-adjusted based on patient case mix variables. While the goal should be to have zero UTIs, the challenge in achieving this goal will vary widely by patient depending upon the clinical condition of the patient—including if the patient is bed-bound due to the clinical condition.

Panel Member #6: Although not originally included in the model, at the suggestion of the TEP, further analysis was performed by suggested clinical factors, such as Bed mobility, transfer, walk in room, walk in corridor, and toilet use. Although there was an increase noted in the measure score noted across all of these factors, the mean proportion did not increase across score deciles.

Similarly, social risk factors were not included in the model due to a lack of evidence but were assessed by race-ethnicity, extreme age, gender, and Medicaid eligibility. Again, while differences were noted, the c-statistic was most prominent at 0.51 for age greater than 85. This indicated weak model performance if age is included.

For cost/resource use measures ONLY:

Panel Member #2: Not applicable.

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):**

VALIDITY: TESTING 2b1

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 2b1.2

    Panel Member #1: Change in performance in the following tests would not necessarily tell us whether a measure is valid nor invalid: 1) Stability analysis, 2) change in performance across years.

    Re the test "Confidence interval analysis": The following analysis is questionable as to whether this tells us something illuminating about validity: "high-performing facilities should have scores that are significantly below average, and scores of low-performing facilities should be significantly above average".

    "Performance Measure Score

    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 [p10]

    Variation by state: We examined whether variation in scores on this measure was substantially attributable to state-by-state differences. [p11]

    Seasonality: Another potential threat to the validity of a quality measure is seasonal variation. [p11]

Stability analysis: We examined the extent to which relative facility rank changed on this quality measure from Quarter 2 to Quarter 3, 2018. [p11]

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. [p11]

Average change in performance across years: … difference in performance scores … across years to assess how updates to the … Instrument 3.0 User's Manual pertaining to item I2300 – UTI … may have changed provider scores from year to year. [p11]

Face validity

RTI convened a TEP on May 23, 2019 to obtain feedback from providers and various stakeholders about the face validity of NQF #0684. TEP members discussed the current measure specifications, potential risk adjustment factors and the effectiveness of the measure in capturing quality of care, to determine the face validity of the measure as it is currently specified (RTI International, 2019)." [p12]

Methods used are appropriate. Extensive testing of measure correlation with other quality measures used to test related parameters in the same facilities, as well as testing of impact of season, state, stability, performance across years, in addition to face validity by a Technical Expert Panel. Generally reasonable, although the methods do not provide an empirical link between the outcome measure of UTI rate and any other process measures of quality of care. The interpretation of differences as having to do with differences in quality of care rests on face validity as judged by TEPs in the various measure submission cycles.

Panel Member #5: The results from the two tables—one identify results from 2003 and one from 2013—show that the national average is virtually 0, with 0 being the value for percentiles 10 – 90. There is virtually no differentiation among nursing homes on this measure. What is the rationale for its continuance?

Panel Member #6: Validity testing was conducted using correlation with other existing clinically relevant measures, variation by state, variation by season, stability over time, confidence interval analysis, average change in performance over years, and by face validity as determined by a convened TEP. Notably, 28% of facilities changed by more than 3 deiles.

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

Submission document: Testing attachment, section 2b1.3, 2b14

Panel Member #1: Some tests were questionable (discussed in Q21 above: Stability analysis, change in performance across years, confidence interval analysis) and the results in general yielded modest results and some poor results. The poorest result was the correlation of measures. The correlation between this measure and two others were pretty weak and in turn do not do a good job of evidencing validity. Also, while I question whether the stability analysis inherently tells us about validity, the finding that more than a quarter (28.4%) of facilities moved more than three deciles in ratings in just one quarter suggests instability vs. stability. The following tests produced positive modest results: Variation by state, Confidence interval analysis & Average change in performance across years.

"Performance Measure Score

Correlation with related quality measures: Correlation between the facility's percentile rank on [this measure] and #0686 (Percent of Residents Who Have/Had a Catheter Inserted and Left in Their Bladder (Long Stay)) and found a positive but weak (ρ = 0.11, and statistically significant (ρ < 0.001)) correlation. orrelation between the facility's percentile rank on [this measure] and NQF #0685 (Percent of Low Risk Residents Who Lose Control of their Bowels or Bladder (Long Stay)) and found a positive (ρ=0.03) and statistically significant (p<0.001) relationship. [p12]

Variation by state: The proportion of variance in this measure explained by the state in which facilities are located is 1.4% ($p < 0.001$). The interquartile range of state-level scores is 4.2% [p12]

Seasonality: … results are presented in ***Figure 1***. … no evidence of seasonal variation. [p12]

Stability analysis: ***Figure 2*** illustrates the changes in facility rank by quality measure score from Quarter 2, 2018, to Quarter 3, 2018. Most (48.5%) facilities are in the same decile in both quarters. Shifts of more than 3 deciles were less common, occurring for approximately 28.4% of facilities. [p13]

Confidence interval analysis: ***Table 2*** shows the proportions of facilities that scored significantly higher or lower. … Overall, 39.5% of facilities were significantly different from the national mean. [p14]

Average change in performance across years: ***Table 3*** presents the changes … . On average, provider scores changed by less than 0.01 percentage points [p15]

Face validity

The majority of TEP members explicitly affirmed the face validity. The TEP supported continued public reporting of the measure." [p16]

Panel Member #2: Methods used are appropriate. No concerns.

Panel Member #3: Demonstrated stability with minimal impact of season, state, with reasonable stability over time and adequate face validity.

Panel Member #5: The results from the two tables—one identify results from 2003 and one from 2013—show that the national average is virtually 0, with 0 being the value for percentiles 10 – 90. There is virtually no differentiation among nursing homes on this measure. What is the rationale for its continuance?

Panel Member #6: Correlation with other measures of quality of care at long term facilities was positive but weak at 0.11. There was no evidence of seasonality. Score stability indicated that almost half of the facilities were in the same decile in two successive quarters. Confidence intervals by difference from the mean did demonstrate and effect of size when analyzed by decile of size. No substantial change in performance was detected in four one-year periods by decile.

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)

    Panel Member #1: Q21 above has my comments in this regard.

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.**

Panel Member #1: As stated in Q22 above: Some tests were questionable (discussed in Q21 above: Stability analysis, change in performance across years, confidence interval analysis) and the results in general yielded modest results and some poor results. The poorest result was the correlation of measures. The correlation between this measure and two others were pretty weak and in turn do not do a good job of evidencing validity. Also, while I question whether the stability analysis inherently tells us about validity, the finding that more than a quarter (28.4%) of facilities moved more than three deciles in ratings in just one quarter suggests instability vs. stability. The following tests produced positive modest results: Variation by state, Confidence interval analysis & Average change in performance across years.

Panel Member #2: Based on the amount of testing and the methods employed, I think that the measure is valid.

Panel Member #3: I remain concerned that the absence of adequate (or any) risk adjustment has not adequately addressed this threat to the validity of the measure.

Panel Member #4: The correlations with other measures reflecting somewhat similar quality concepts are statistically significant, if not strong.

Panel Member #5: Given that lack of risk adjustment and virtually no differentiation among SNFs for this measure, why are we keeping this measure in the inventory?

Panel Member #6: The measure developers presented not only face validity but tested a number of potential threats to validity, including social risk factors and TEP-suggested clinical factors, and, despite these analyses, the measure score demonstrated at least moderate validity and could be considered for high validity..

FOR COMPOSITE MEASURES ONLY: Empirical analyses to support composite construction

Panel Member #2: Not applicable.

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

☐ **High**

☐ **Moderate**

☐ **Low**

☐ **Insufficient**

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

ADDITIONAL RECOMMENDATIONS

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

Panel Member #1: Regarding concerns expressed about clarify of measure specifications (see Q2 above): Suggest further specificity be provided regarding the denominator definition.

Panel Member #4: This is a difficult measure to evaluate, as it has already been endorsed by NQF twice, and the measure developer is clearly very experienced and has done generally reasonable things in terms of testing reliability and validity. The absence of any formal testing of risk adjustment models, and therefore the absence of any risk adjustment, is troubling, as there are clearly variables that make a difference in the likelihood of UTI that are ignored in calculating facility-level scores. The measurement "signal" for any facility, therefore, is some mix of actual quality of care and unadjusted case mix, leading to inflated estimates of signal-to-noise ratios and reliability.

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.

All the data necessary to calculate this measure are found within MDS 3.0 which is collected by all Medicare approved nursing homes.

Questions for the Committee:

- Are the required data elements routinely generated and used during care delivery?
- Are the required data elements available in electronic form, e.g., EHR or other electronic sources?

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

Committee Pre-evaluation Comments:
Criteria 3: Feasibility

3. Feasibility

Comments:

**No concerns

**Not my area of expertise, but seems to be readily available data

**No concerns

**Seems very feasible with little burden on the provider since it uses info from regular assessments done for nursing home patients. It is not clear how they determine if the patient had a UTI in the prior 30 days - these minimum data sets include treatments, which presumably would include UTI treatment.

**high

**Data already collected using MDS 3.0

**MDS has elements routinely collected -not sure about eCQMs

**none

**many nursing homes do not use EHR

**no

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 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 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 Quality Measure (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. 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 scores on this quality measure via the Nursing Home Compare website

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

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

Six out of 10 TEP members rated this measure as "very important" (rating it a 4 or 5), noting that it is a critical health outcome and that this quality measure promotes accurate diagnosis and tracking of UTIs as well as the appropriate use of antibiotics.

A majority of TEP members also noted that they use this QM to track facility performance and address concerns, including reviewing residents that trigger the QM, ensuring proper documentation, evaluating staff understanding of instructions in the Long-Term Care Facility Resident Assessment Instrument 3.0 User's Manual, addressing clinical practices, and comparing their facility to state and national benchmarks. In particular, with respect to the consumer perspective, one TEP member stated that this QM "provide[s] good information for follow up questions to [the] facility."

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 recent (October 2017) changes in UTI coding guidelines in the Long-Term Care Facility Resident Assessment Instrument 3.0 User's Manual. One TEP member suggested that mobility limitations be considered as a potential risk adjuster because "The increased pooling of urine/urinary stasis, the higher level at which the urethra sits in the prone patient, and the weaker extruder muscles that can lead to incomplete emptying all may increase risk of UTI in the immobile patient." Based on this suggestion, RTI conducted testing of several MDS data elements assessing function. Results (presented in 2b3.3a. in the Testing Attachment) demonstrated that there were not strong relationships between measure performance and functional data elements. Therefore, we did not proceed with further risk adjustment testing.

Eight out of 10 TEP members reported no evidence or rationale for modifying the measure specifications to include exclusion criteria or to use a statistical risk model or stratification to adjust for resident social or clinical risk factors. TEP members noted that risk adjustment could mask important disparities, rather than help providers and consumers better assess facility performance.

A minority of TEP members questioned whether the UTI measure reflected quality of care in nursing homes and whether it could be used for improvement. They were concerned that nursing home staff may not apply evidence-based criteria to diagnose UTIs as intended and that some residents may develop UTIs despite a nursing home's best efforts to prevent them. However, other TEP members countered this argument, noting that the quality measure "allows for better accountability on the part of those diagnosing UTIs" (RTI, 2019).

Additional Feedback:

No additional feedback was provided by the developer.

***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?

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

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):

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. After an early increase in mean and median scores, the national facility-level mean and median scores have trended steadily downward since the implementation of the MDS 3.0 measure, indicating a general improvement in performance over time. The mean score for this measure was 7.5% in Quarter 1, 2011, and the median score was 6.5%. In Quarter 3, 2018, the mean and median scores were 2.8% and 1.9%, respectively.

All United States Medicare/Medicaid certified nursing homes with eligible long-stay residents. In Quarter 3, 2018, there were 15,299 eligible facilities containing 1,104,673 residents eligible for inclusion in the measure

(with both prior and target assessments); 14,520 facilities (95.0%) containing 1,096,778 residents (99.3%) had sufficient sample size (20 or more long-stay residents included in the denominator) to report on this measure.

Unexpected findings (positive or negative) during implementation

The developer did not report unexpected findings.

Potential harms

The developer did not report potential harms.

Additional Feedback:

The developer did not report additional feedback.

Questions for the Committee:

- 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:

**Used in public reporting and judged as an important measure

**The measure is publicly reported and feedback provided as per the documents provided

**Previously approved twioce and in use by other bodies - and with apparent benefit.

**The measure is used on Nursing Home Compare for long-stay residents. Fairly easy for the public to find & I think understandable. The documents indicate the measure is used to provide feedback; developer convened a TEP to see if any updates were needed.

**yes

**Publically reported and used in accountability program Nursing Home Quality Initiative (NHQI)

**yes

**none - are we measuring thr right thing if there has been little change

**n/a

**Yes

4b1. Usability – Improvement

Comments:

**Since it is an outcome measure, its use for improvement is clear. No concerns for harm

**There should be minimal harm. If the metric was more valid and better accounted for patient characteristics, duration of catheter use, it would be an important metric to assess. Perhaps it should be limited to catheters in place for a shorter period of time where processes are more closely tied to outcome and can be modifiable?

**Curret data show usable and without harms.

**Don't see any harm or unintended consequences but wonder if this measure is making a difference when provider scores changed by less than 0.01 percentage points 2014-18

**none noted

** High – general improvement shown over time, continue to promote evidence based practices at nursing homes

**none reported

**none

**overtreatment, inadequate sample techniques, overuse of indwelling catheters

**None identified

Criterion 5: Related and Competing Measures

Related or competing measures

The developer listed two related measures:
0138 : National Healthcare Safety Network (NHSN) Catheter-associated Urinary Tract Infection (CAUTI) Outcome Measure
0281 : Urinary Tract Infection Admission Rate (PQI 12)

One measure with a similar focus that is not NQF-endorsed is also noted – Urinary tract infection (UTI) admission rate (area-level): rate per 100,000 population (AHRQ).

Harmonization

0138: National Healthcare Safety Network (NHSN) Catheter-associated Urinary Tract Infection (CAUTI) provides the Standardized Infection Ratio (SIR) of healthcare-associated, catheter-associated urinary tract infections (UTI) among patients in bedded inpatient care locations, except level II or level III neonatal intensive care units (NICU). This includes acute care general hospitals, long-term acute care hospitals, rehabilitation hospitals, oncology hospitals, and behavior health hospitals. Although related to UTIs, CAUTIs reflect a distinct issue that may require different clinical intervention; as such, providers' efforts to prevent UTIs and CAUTIs may vary. For example, CAUTI prevention includes reducing the number of unnecessary indwelling catheters inserted, removing indwelling catheters at the earliest possible time, securing catheters to the patient's leg to avoid bladder and urethral trauma, keeping the urine collection bag below the level of the bladder, and utilizing aseptic technique for urinary catheter insertion. It may be challenging to measure CAUTI in nursing homes due to concerns about the availability of onsite laboratory testing; subsequently, reportability of a nursing home CAUTI measure may be substantially diminished.

0281: Urinary Tract Infection Admission Rate (PQI 12) reports the rate of admissions with a principal diagnosis of urinary tract infection per 100,000 population, ages 18 years and older. Patients with kidney or urinary tract disorder admissions, other indications of immunocompromised state admissions, obstetric admissions, and transfers from other institutions are excluded. Presence of a urinary tract infection is based on a principal diagnosis code (ICD-9) for UTI. UTIs in the adult population may generally be treated in ambulatory/outpatient care settings. However, when treatment is inadequate or delayed, patients may develop more severe clinical infections; as a result, they may be more likely to present at an emergency department and, subsequently, require inpatient admission. Therefore, access to sufficient outpatient care may be key to reducing urinary tract infection admissions. Although NQF #0281 and Percent of Residents With a Urinary Tract Infection (Long Stay) (NQF #0684) both capture UTI rates, they are intended for use in disparate populations (adults utilizing inpatient acute care vs. long-stay nursing home residents).

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

5. Related and Competing

Comments:

**Will speak to concerns about overlap with 0686 in its review

**The NHSN metric, although not ideas is more closely aligned with clinical UTIs.

**0138/0281 are in same area, and these seem aligned and nonredundant.

**no

**0138 and 0281

**0138 : National Healthcare Safety Network (NHSN) Catheter-associated Urinary Tract Infection (CAUTI) Outcome Measure – different issue and prevention effort required; 0281 : Urinary Tract Infection Admission Rate (PQI 12) – different population measured (inpatient acute care patients vs. long-star nursing home residents). Need to harmonize?

**there are related but not in NH

**yes with cauti and one other

**not that i know of

**None identified

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.**

Brief Measure Information

NQF #: 0684

Corresponding Measures:

De.2. Measure Title: Percent of Residents with a Urinary Tract Infection (Long Stay)

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

De.3. Brief Description of Measure: This measure reports the percentage of long-stay residents in a nursing home who have a urinary tract infection in the 30 days prior to the target assessment. 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:

Nursing home residents frequently develop infections (Herzig et al., 2017) and, among these, urinary tract infections (UTIs) are the most common (Herzig et al., 2017; Smith et al., 2018). Symptoms of UTIs include fever, painful or difficult urination, increased frequency and urgency of urination, blood in the urine, low abdominal or flank pain or tenderness, and deterioration in mental status (such as increased confusion). Thus, improving quality of care in genitourinary health domains, including reducing UTIs, may improve quality of life for long-stay nursing home residents in part by decreasing associated signs and symptoms, especially those that cause pain or discomfort.

UTIs can also lead to serious corollary outcomes and complications, such as sepsis, hospitalization, emergency department use, delirium, or death (Castle et al., 2017; Wolff, 2016). For example, some residents who develop UTIs develop blood infections, and 10 percent of these patients die within one week (Gould et al., 2009). UTIs are also a leading cause of avoidable hospitalizations (Nelson & Flynn, 2015; Wolff et al., 2017). Additionally, an increasingly prevalent corollary outcome is antimicrobial resistance in UTIs. Antimicrobial resistance proliferates through inappropriate widespread antibiotic treatment of asymptomatic bacteriuria (often misdiagnosed as UTI), which has no benefits for patients and can cause potential harm such as adverse drug reactions, drug resistance, and subsequent antibiotic-resistant infections (Cooper et al., 2019; Crnich & Drinka, 2014). Therefore, reducing UTIs, as well as improving diagnosis and treatment of UTIs, may improve long-stay nursing home residents' health outcomes.

Gaps in Performance in Nursing Homes:

The UTI quality measure is the only measure in the current measure set that addresses infections and, thereby, is a very important indicator of how facilities prevent and manage infections. UTI rates have been linked to modifiable nursing home factors, such as nurse staffing levels and mix, administering medication on time and documenting nursing care, and education in catheterization protocols and infection prevention measures (Meddings et al., 2017, Trautner et al., 2017; Nelson & Flynn, 2015). This quality measure should encourage nursing homes to direct resources to these care domains.

Since many UTIs are related to catheters, this quality measure provides an additional incentive for the facility to monitor its catheter use (RTI, 2019b). Nursing homes also vary in their adoption of evidence-based practices for UTI prevention. Practices such as adhering to clinical guidelines, keeping the perineal area clean, ensuring hand hygiene, improving management of urinary incontinence, and implementing hydration regimens have been shown to be effective in preventing UTIs (Montoya et al, 2016; Meddings et al., 2017; Wolff et al., 2017). Thus, this quality measure may also promote wider use of these practices.

Gaps in Performance among Specific Groups of Nursing Home Residents:

There is some evidence in the literature of empirical relationships between patient characteristics and UTI rates:

1. One study identified a positive relationship between female gender and UTI rate (Gucwa et al., 2016).

2. White race has been found to be a predictor of UTIs (Hefele et al., 2017; Castle et al., 2017).

3. Older age is associated with higher rates of UTIs (Castle et al., 2017).

4. Higher rates of Medicaid coverage in a facility are negatively associated with UTI rates among non-catheterized residents and positively associated with UTI rates among catheterized residents (Castle et al., 2017).

Overall, these findings are consistent with RTI's analysis, which found that individuals who were of non-Hispanic white race/ethnicity, older, and female were slightly more likely than their counterparts to have a UTI. In addition, our analyses also demonstrated that individuals who were not eligible for Medicaid were slightly more likely than individuals who were eligible for Medicaid to have a UTI.

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 (MDS) 3.0. In the pre-TEP survey, six out of 10 TEP members rated the UTI 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); has a demonstrated high-impact aspect of healthcare (e.g., affects large numbers); has external evidence of importance, such as consensus standards; and/or has evidence of disparities for the quality domain (RTI, 2019b).

The TEP noted that it was important to maintain public reporting of the UTI quality measure as it reflects a critical health outcome that warrants continued attention, citing significant negative outcomes of UTIs on nursing home residents' function, quality of life, and socialization. The TEP believed this quality measure was important for promoting quality improvement for the following reasons:

1. The UTI quality measure has an important role in promoting accountability and tracking urinary tract infection, allowing facilities to "identify patterns and implement solutions."

2. The presence of this quality measure keeps providers' efforts focused on improving accurate diagnosis and can also lead to improved antibiotic prescribing and antibiotic stewardship.

3. The TEP indicated the UTI quality measure encouraged investment in education and training for clinical and direct care staff.

Castle, N., Engberg, J. B., Wagner, L. M., & Handler, S. (2017). Resident and facility factors associated with the incidence of urinary tract infections identified in the nursing home Minimum Data Set. Journal of Applied Gerontology, 36(2), 173-194.

Cooper, D., McFarland, M., Petrilli, F., & Shells, C. (2019). Reducing inappropriate antibiotics for urinary tract infections in long-term care: A replication study. Journal of Nursing Care Quality, 34(1), 16-21.

Crnich, C. J. & Drinka, P. (2014). Improving the management of urinary tract Infections in nursing homes: It´s time to stop the tail from wagging the dog." Annals of Long Term Care: Clinical Care and Aging, 22(9), 32-36.

Gould CV, Umscheid CA, Agarwal R, Kuntz G, Pegues DA; the Healthcare Infection Control Practices Advisory Committee. Guideline for prevention of catheter-associated urinary tract infections 2009. Atlanta: Centers for Disease Control and Prevention, 2009. Available from

https://www.cdc.gov/infectioncontrol/pdf/guidelines/cauti-guidelines-H.pdf.

Herzig, C., Dick, A., Castle, N., & Stone, P. (2016). Policies and practices to reduce urinary tract infections in nursing homes [Abstract]. American Journal of Infection Control 44, S12.

Meddings, J., Saint, S., Krein, S. L., Gaies, E., Reichert, H., Hickner, A.,…Mody, L. (2017). Systematic review of interventions to reduce urinary tract infection in nursing home residents. Journal of Hospital Medicine, 12(5), 356

Montoya, A., Cassone, M., & Mody, L. (2016). Infections in nursing homes: Epidemiology and prevention programs. Clinics in Geriatric Medicine, 32(3), 585-607.

Nelson, S. T., & Flynn, L. (2015). Relationship between missed care and urinary tract infections in nursing homes. Geriatric Nursing, 36(2), 126-130.

RTI analysis of MDS 3.0 episode files for Quarter 1, 2011–Quarter 3, 2018 (programming reference: KH46\hf15_request_684_31_32.log)

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.

Smith, S. N., Greene, M. T., Mody, L., Banaszak-Holl, J., Petersen, L. D., & Meddings, J. (2018). Evaluation of the association between Nursing Home Survey on Patient Safety culture (NHSOPS) measures and catheter-associated urinary tract infections: Results of a national collaborative. BMJ Quality & Safety, 27(6), 464-473.

Trautner, B. W., Greene, M. T., Krein, S. L., Wald, H. L., Saint, S., Rolle, A. J.,...Mody, L. (2017). Infection prevention and antimicrobial stewardship knowledge for selected infections among nursing home personnel. Infection Control & Hospital Epidemiology, 38(1), 83-88.

Wolff, M. L., et al. (2016). An innovative quality assurance activity to reduce urinary tract infection rates in a green house skilled nursing setting." Annals of Long Term Care, 24(10), 17-20.

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 a urinary tract infection within the last 30 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) and who do not meet the exclusion criteria.

S.8. Denominator Exclusions: If the target assessment is an OBRA Admission Assessment, PPS 5-Day Assessment, or PPS Readmission/Return Assessment, the resident is excluded from the denominator for this quality measure. A resident is also excluded if the target assessment indicates that data is missing for the data element assessing urinary tract infection in the last 30 days.

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:** Nov 10, 2014

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

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

De.4. IF PAIRED/GROUPED, what is the reason this measure must be reported with other measures to appropriately interpret results? Not applicable; this measure is not paired/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_0684_UTI_Evidence_Form_10-31-19_508.docx

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

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

Yes

1a. Evidence (subcriterion 1a)

NATIONAL QUALITY FORUM—Evidence (subcriterion 1a)

Measure Number (*if previously endorsed*)**:** 0684

Measure Title: Percent of Residents with Urinary Tract Infection

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: Urinary Tract Infection

☐ 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 with a urinary tract infection (UTI). UTIs are important to address both because of their impacts on health as an intermediate outcome (such as unpleasant symptoms) and because they can lead to serious corollary outcomes and complications, such as sepsis, hospitalization, emergency department use, delirium, or death.

Additionally, an increasingly prevalent corollary outcome is antimicrobial resistance in UTIs. In one study, Kistler et al. (2017) found MDROs in 12% of bacterial UTI cultures across 31 nursing homes. Antimicrobial resistance proliferates through inappropriate widespread antibiotic treatment of asymptomatic bacteriuria (often misdiagnosed as UTI), which has no benefits for patients (Cooper et al., 2019; Crninch & Drinka, 2014). The growing prevalence of antibiotic resistance increases the importance of preventing UTIs in coordination with treatment efforts and only prescribing antibiotics when necessary and clinically appropriate.

Nursing home structural characteristics and care processes can impact the quality of care that facilities provide to residents (Mukamel et al., 2008) and may therefore ultimately impact the health outcomes of residents, including UTIs. ***Figure 1*** below illustrates the key structures and outcomes that are associated with lower rates of UTI. 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

Nursing home characteristics associated with higher UTI incidence include chain membership (i.e., a type of ownership structure where 2 or more facilities share the same owner) and occupancy rate (Castle & Anderson, 2011). Nurse staffing levels have also been found to be associated with UTI rates (Castle & Anderson, 2011; Castle et al., 2017; Hyer et al., 2011). In addition to nurse staffing levels, organizational efforts to educate staff in catheter protocols and infection prevention have successfully reduced UTI rates (Meddings et al., 2017; Trautner et al., 2017). One study found training for staff from the Association for Professionals in Infection Control and Epidemiology to be associated with lower rates of UTIs (Herzig et al., 2016).

Evidence for link between processes and quality of care outcomes

Nursing home characteristics, infrastructure, and resources may affect key processes known to influence UTI outcomes. As described in the text that follows, these key processes are adherence to clinical guidelines, appropriate antibiotic treatment, appropriate assessments of and use of catheters, and best practices in hygiene and hydration.

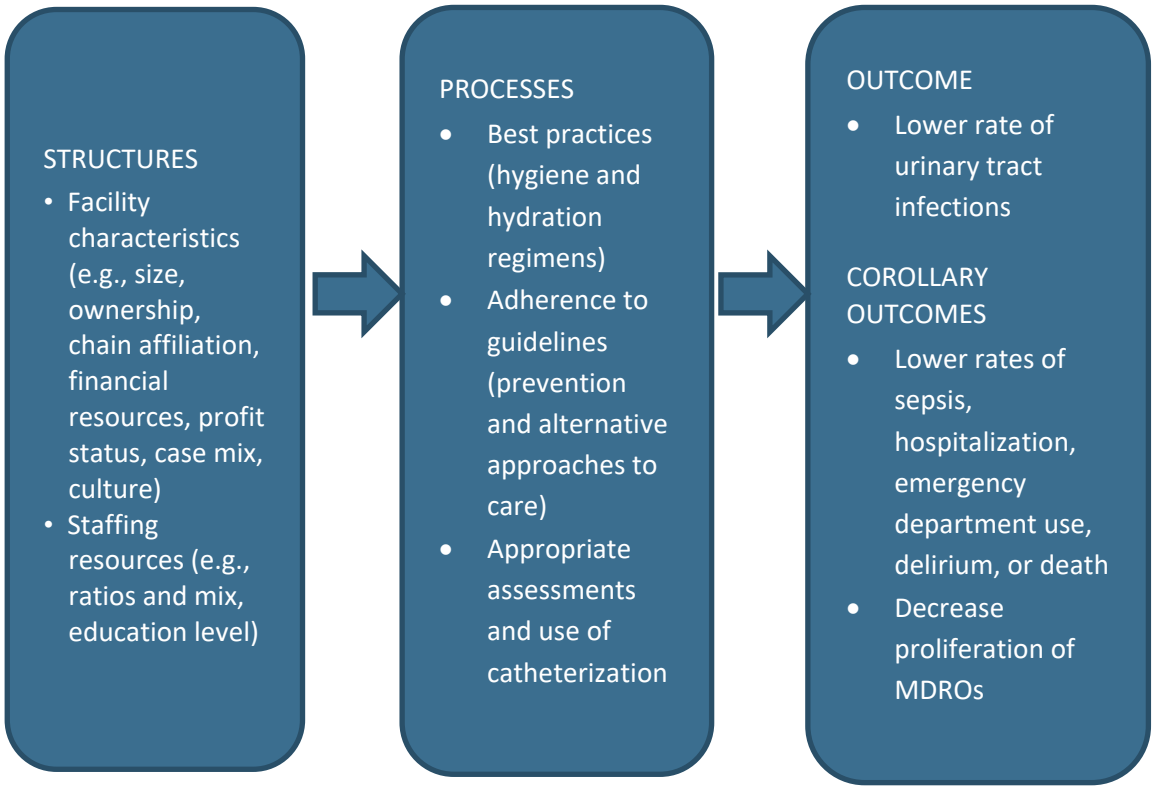
First, implementation of clinical guidelines for UTI diagnostic processes and treatment, including the prescription of antibiotics, reduces adverse outcomes (Cooper et al., 2019; McMaughan et al., 2016; Nicolle, 2016). To relieve the burning pain and urgent need to urinate associated with UTI and prevent complications, UTIs are usually successfully treated with antibiotics and other medications (Detweiler et al., 2015; Nicolle, 2009; Saint et al., 2006; Gradwohl et al., 2016). In fact, one study estimated that antibiotic therapy shortens the duration of symptoms and likely cures more than 90 percent of infections (Nicolle et al., 2006). Clinical guidelines can help nursing homes identify the cause of a UTI and implement appropriate treatment sooner, which can prevent the infection from becoming more serious or causing serious complications, such as sepsis and death (Smith et al., 2018). Consistent surveillance of residents is also important in UTI prevention, with missing activities such as administering medication on time and documenting nursing care associated with higher rates of UTIs (Nelson & Flynn, 2015).

Second, one commonly cited evidence-based clinical practice for preventing UTIs and reducing UTI rates is appropriate use of catheterization. Generally, the use of catheters is associated with urinary tract infections, with one study finding that 3% to 7% of residents with indwelling catheters develop UTIs (Gucwa et al., 2016). Moreover, the odds of a resident having a multi-drug resistant organism (MDRO) UTI significantly increased with the presence of catheterization (Gucwa et al., 2016). Therefore,

encouraging only appropriate use of urinary catheterization can help to lower the rate of UTIs and prevent the corollary adverse outcome of MDRO proliferation.

Third, evidence suggests that there are broader clinical practices nursing homes can implement, beyond appropriate catheterization, that are effective in preventing UTIs. These processes include keeping the perineal area clean (Wolff et al., 2017), ensuring hand hygiene (Montoya et al., 2016; Meddings et al., 2017), improving management of urinary incontinence (Meddings et al., 2017), and implementing hydration regimens (Wolff et al., 2017).

Figure 1: Role of Nursing Home Structures and Processes in Rate of Urinary Tract Infections



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.

Castle, N., Engberg, J. B., Wagner, L. M., & Handler, S. (2017). Resident and facility factors associated with the incidence of urinary tract infections identified in the nursing home Minimum Data Set. *Journal of Applied Gerontology, 36*(2), 173-194.

Cooper, D., McFarland, M., Petrilli, F., & Shells, C. (2019). Reducing inappropriate antibiotics for urinary tract infections in long-term care: A replication study. *Journal of Nursing Care Quality, 34*(1), 16-21.

Crnich, C. J. & Drinka, P. (2014). Improving the management of urinary tract Infections in nursing homes: It's time to stop the tail from wagging the dog." *Annals of Long Term Care: Clinical Care and Aging, 22*(9), 32-36.

Detweiler, K., Mayers, D., & Fletcher, S. G. (2015). Bacteriuria and urinary tract infections in the elderly. *Urologic Clinics of North America, 42*(4), 561-568.

Gradwohl S.E., Bettcher, C.M., Chenoweth C.E., Van Harrison R., & Zoschnick L.B. (2016). Michigan Medicine: Urinary tract infection [PDF]. Retrieved from http://www.med.umich.edu/1info/FHP/practiceguides/uti/uti.pdf

Gucwa, A. L., Dolar, V., Ye, C., & Epstein, S. (2016). Correlations between quality ratings of skilled nursing facilities and multidrug-resistant urinary tract infections. *American Journal of Infection Control, 44*(11), 1256-1260.

Herzig, C., Dick, A., Castle, N., & Stone, P. (2016). Policies and practices to reduce urinary tract infections in nursing homes [Abstract]. *American Journal of Infection Control 44*, S12.

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.

Meddings, J., Saint, S., Krein, S. L., Gaies, E., Reichert, H., Hickner, A.,…Mody, L. (2017). Systematic review of interventions to reduce urinary tract infection in nursing home residents. *Journal of Hospital Medicine*, *12*(5), 356.

Montoya, A., Cassone, M., & Mody, L. (2016). Infections in nursing homes: Epidemiology and prevention programs. *Clinics in Geriatric Medicine, 32*(3), 585-607.

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.

Nelson, S. T., & Flynn, L. (2015). Relationship between missed care and urinary tract infections in nursing homes. *Geriatric Nursing, 36*(2), 126-130.

Nicolle L., Anderson P.A.M., Conly J., Mainprize T.C., Meuser J, Nickel J.C.,…Zhanel G.G. (2006) Uncomplicated urinary tract infections in women: current practice and the effect of antibiotic resistance on empiric treatment. *Canadian Family Physician, 52,* 612-618.

Nicolle L. E. (2009). Urinary tract infections in the elderly. *Clinics in Geriatric Medicine, 25*(3), 423-36.

Nicolle, L. E. (2016). Urinary tract infections in the older adult. *Clinics in Geriatric Medicine, 32*(3), 523-538.

Saint S., Kaufmann S.R., Rogers M.A.M., Baker, P.D., Boyko, E.J ,& Lipsky, B.A. (2006) Risk factors for nosocomial urinary tract-related bacteremia: A case control study. *American Journal of Infection Control*, *34*(7),401-7.

Smith, S. N., Greene, M. T., Mody, L., Banaszak-Holl, J., Petersen, L. D., & Meddings, J. (2018). Evaluation of the association between Nursing Home Survey on Patient Safety culture (NHSOPS) measures and catheter-associated urinary tract infections: Results of a national collaborative. *BMJ Quality & Safety, 27*(6), 464-473.

Trautner, B. W., Greene, M. T., Krein, S. L., Wald, H. L., Saint, S., Rolle, A. J.,...Mody, L. (2017). Infection prevention and antimicrobial stewardship knowledge for selected infections among nursing home personnel. *Infection Control & Hospital Epidemiology, 38*(1), 83-88.

Wolff, M. L., et al. (2016). An innovative quality assurance activity to reduce urinary tract infection rates in a green gouse skilled nursing setting." *Annals of Long Term Care*, *24*(10), 17-20.

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.**

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. Systematic Review of Interventions to Reduce Urinary Tract Infection in Nursing Home Residents (2017)

Ref: Meddings, J., Saint, S., Krein, S. L., Gaies, E., Reichert, H., Hickner, A.,…Mody, L. (2017). Systematic review of interventions to reduce urinary tract infection in nursing home residents. *Journal of Hospital Medicine*, *12*(5), 356.

Assessment of Interventions for Patients Regardless of Urinary Catheter Status

- Hand hygiene (R)
- Encourage fluid intake/hydration to reduce infection (NR)
- Improve general patient hygiene to reduce infection (R)
- Cranberry product as prophylaxis (NR)
- Vitamin/mineral supplement as UTI prophylaxis (NR)
- Treatment of atrophic vaginitis as UTI prophylaxis (R)
- Interventions to improve management of urinary incontinence (R)
- Implementation of effective infection control program (R)

Table 1. Recommendations for Interventions

Assessment	Definition
R	**Interventions that have some evidence of benefit (not always from controlled intervention studies), at least for certain populations and settings.**
NR	**Interventions that are not recommended based on available evidence or rationale.**

2. Clinical Practice Guideline for the Evaluation of Fever and Infection in Older Adult Residents of Long-Term Care Facilities (2009)

Ref: High K.P., Bradley S.F., Gravenstein D., Mehr D.R., Quagliarello V.J., Richards C., & Yoshikawa T.T. (2009). Clinical practice guideline for the evaluation of fever and infection in older adult residents of long-term care facilities. *Clinical Infectious Diseases*,*48*,149-171. https://academic.oup.com/cid/article/48/2/149/304388

Guidelines (See Table 2 for grading of evidence)

Urinalysis and Urine Culture

12. Urinalysis and urine cultures should not be performed for asymptomatic residents (A-I).

13. In noncatheterized residents, the diagnostic laboratory evaluation of suspected UTI should be reserved for those with acute onset of UTI-associated symptoms and signs (e.g., fever, dysuria, gross hematuria, new or worsening urinary incontinence, and/or suspected bacteremia) (A-II).

14. In residents with long-term indwelling urethral catheters, evaluation is indicated if there is suspected urosepsis (i.e., fever, shaking chills, hypotension, or delirium), especially in the context of recent catheter obstruction or change (A-II).

15. Appropriately collected urine specimens include a midstream or clean-catch specimen obtained from elderly men who are cooperative and functionally capable; however, it is often necessary to use a freshly applied, clean condom external collection system, with frequent monitoring of the urine bag (B-II). Specimen collection from women will often require an in and-out catheterization (B-III).

16. Residents with long-term indwelling urethral catheters and suspected urosepsis should have catheters changed prior to specimen collection and institution of antibiotic therapy (A-II).

17. The minimum laboratory evaluation for suspected UTI should include urinalysis for determination of leukocyte esterase and nitrite level by use of a dipstick and a microscopic examination for WBCs (B-II). If pyuria (110 WBCs/high-power field or a positive leukocyte esterase or nitrite test is present on dipstick, only then should a urine culture (with antimicrobial susceptibility testing) be ordered (B-III).

18. If urosepsis is suspected, urine and paired blood specimens should be obtained, if feasible, for culture and antimicrobial susceptibility testing, and a Gram stain of uncentrifuged urine should be requested (B-III).

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.

3. Urinary Tract Infections in Long-Term Care Facilities (2001)

Ref: Nicolle, L.E., SHEA Long-Term Care Committee. Urinary tract infections in long-term care facilities [PDF]. (2001). *Infection Control and Hospital Epidemiology, 22*(3), 167-175. Retrieved from http://www.shea-online.org/images/guidelines/UTIs_in_LTCF_2001.pdf

Guidelines (See Table 3 for grading of evidence)

Surveillance

1. Surveillance for endemic asymptomatic bacteriuria in LTCFs should not be undertaken. Category BII.
    a. Surveillance may be appropriate when an outbreak with a potential uropathogen is suspected. *Category BIII.*
2. Surveillance for symptomatic infection may be undertaken, depending on institutional priorities and infection control resources. *Category BII.*
    a. Rates should be reported as per 1,000 patient-days or per 1,000 catheterized-patient–days. *Category BIII.*
3. Standard diagnostic criteria should be used for the identification of symptomatic UTI. Limitations in the ability to make a specific diagnosis of symptomatic UTI should be acknowledged in these criteria. *Category BII.*

Prevention

1. Nutritional status and care of comorbid illnesses should be optimized for patients in LTCFs. *Category CIII.*
    a. There is insufficient evidence for a recommendation for routine use of estrogen therapy to prevent UTI in women. No recommendation.

2. Routine screening for UTI by urinalysis or urine culture is not recommended for LTCF residents. *Category AI.*

3. Condom catheters should be used to manage incontinence in men only where the benefits to the patient outweigh potential risks. *Category AII.*

   a. Condom catheters for external urinary drainage should be applied and managed appropriately to minimize skin breakdown and ensure unobstructed drainage. *Category AIII.*

   b. Condom-catheter leg bags should be disinfected and dried prior to reuse. *Category CII.*

4. Where clinically appropriate, intermittent catheterization should be used for urinary drainage rather than a chronic indwelling catheter. *Category BII.*

   a. For intermittent catheterization, use a clean technique. *Category AI.*

5. Chronic indwelling urethral catheters should be used only where the benefits outweigh the risks of UTI and its complications. *Category AII.*

   a. Indwelling catheters should be discontinued at the earliest opportunity. *Category AII.*

   b. There is insufficient evidence to make a recommendation for or against routine chronic indwelling urethral catheter changes. No recommendation.

6. Bacteriuric LTCF residents who are to undergo an invasive genitourinary procedure should receive preprocedure antimicrobial prophylaxis. *Category AI.*

Diagnosis

1. A clinical diagnosis of symptomatic UTI should be made only with acute symptoms referable to the genitourinary tract or bacteremia. *Category BII.*

   a. A clinical diagnosis of symptomatic UTI should not be made in the presence of stable, chronic genitourinary symptoms. *Category BIII.*

   b. In residents with clinical deterioration, including fever, no localizing genitourinary findings, and a positive urine culture, a diagnosis of UTI is possible but not definite. *Category AII.*

2. A urine specimen for culture should be obtained prior to therapy from any resident treated for symptomatic UTI. *Category AII.*

   a. A urine specimen with 105 CFU/mL of organisms is consistent with UTI. *Category AI.*

   b. In the presence of acute urinary symptoms, lower quantitative counts may be consistent with the diagnosis of acute UTI. *Category BIII.*

   c. The diagnosis of asymptomatic bacteriuria requires two consecutive urine specimens with $>=10^5$ CFU/mL of the same organism and the absence of symptoms referable to the urinary tract. *Category AII.*

   d. For men and women, a clean-catch voided urine specimen is the preferred method for collection of urine for culture. *Category AII.*

   e. In men using external condom collecting systems, a urine specimen collected from a freshly applied leg bag with $>=10^5$ CFU/mL is consistent with UTI. *Category AII.*

   f. For men or women, where a voided specimen cannot be collected, a urine specimen should be obtained by in-and-out catheterization. A quantitative count $>=10^3$ CFU/mL of a single predominant pathogen from a specimen obtained with appropriate aseptic technique is consistent with infection. *Category BIII.*

3. In asymptomatic patients with chronic indwelling catheters, urine specimens for culture should be obtained aseptically through the catheter port. *Category AII.*

   a. In patients with chronic indwelling catheters and suspected symptomatic UTI, a urine specimen for culture to determine infecting organisms and susceptibilities should be obtained from a freshly inserted chronic indwelling catheter prior to initiating antimicrobial therapy. *Category AI.*

4. The presence or absence of pyuria in a urinalysis specimen should not be used as a criterion to diagnose UTI or to differentiate symptomatic from asymptomatic infection. *Category AII.*

   a. The absence of pyuria makes UTI unlikely. *Category AII.*

Treatment

1. Asymptomatic bacteriuria should not be treated with antimicrobial therapy in LTCF residents. *Category A1.*

2. There is insufficient evidence to recommend nonantimicrobial methods to manage UTI in LTCF residents. No recommendation.

3. The selection of an antimicrobial for treatment of symptomatic UTI should be based on known or suspected infecting organisms, patient tolerance, local formulary, and documented efficacy. *Category AIII.*

   a. The duration of therapy should not exceed 10 to 14 days; shorter courses may be considered for women with minor lower tract symptoms. *Category AII.*

   b. For individuals with chronic indwelling catheters, the duration of therapy should be less than 10 days. *Category BIII.*

   c. For men with relapsing symptomatic infection, 6 weeks of therapy for retreatment may be considered. *Category BI.*

4. When the diagnosis of symptomatic UTI is uncertain, a decision of whether or not to treat with antimicrobials must be made on the basis of clinical assessment. Where antimicrobials are given, ongoing clinical reassessment of presenting signs and symptoms to assess the impact of antimicrobial therapy should be undertaken. *Category BIII.*

   a. Clinical assessment should be undertaken by a physician or appropriately trained designate. *Category AII.*

5. Post-treatment urine cultures to document cure should not be obtained. *Category AII.*

Table 3. Strength of Recommendation and Quality of Evidence

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

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.

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.

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

This is not applicable.

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

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:

Nursing home residents frequently develop infections (Herzig et al., 2017) and, among these, urinary tract infections (UTIs) are the most common (Herzig et al., 2017; Smith et al., 2018). Symptoms of UTIs include fever, painful or difficult urination, increased frequency and urgency of urination, blood in the urine, low abdominal or flank pain or tenderness, and deterioration in mental status (such as increased confusion). Thus, improving quality of care in genitourinary health domains, including reducing UTIs, may improve quality of life for long-stay nursing home residents in part by decreasing associated signs and symptoms, especially those that cause pain or discomfort.

UTIs can also lead to serious corollary outcomes and complications, such as sepsis, hospitalization, emergency department use, delirium, or death (Castle et al., 2017; Wolff, 2016). For example, some residents who develop UTIs develop blood infections, and 10 percent of these patients die within one week (Gould et al., 2009). UTIs are also a leading cause of avoidable hospitalizations (Nelson & Flynn, 2015; Wolff et al., 2017). Additionally, an increasingly prevalent corollary outcome is antimicrobial resistance in UTIs. Antimicrobial resistance proliferates through inappropriate widespread antibiotic treatment of asymptomatic bacteriuria (often misdiagnosed as UTI), which has no benefits for patients and can cause potential harm such as adverse drug reactions, drug resistance, and subsequent antibiotic-resistant infections (Cooper et al., 2019; Crnich & Drinka, 2014). Therefore, reducing UTIs, as well as improving diagnosis and treatment of UTIs, may improve long-stay nursing home residents' health outcomes.

Gaps in Performance in Nursing Homes:

The UTI quality measure is the only measure in the current measure set that addresses infections and, thereby, is a very important indicator of how facilities prevent and manage infections. UTI rates have been linked to modifiable nursing home factors, such as nurse staffing levels and mix, administering medication on time and documenting nursing care, and education in catheterization protocols and infection prevention measures (Meddings et al., 2017, Trautner et al., 2017; Nelson & Flynn, 2015). This quality measure should encourage nursing homes to direct resources to these care domains.

Since many UTIs are related to catheters, this quality measure provides an additional incentive for the facility to monitor its catheter use (RTI, 2019b). Nursing homes also vary in their adoption of evidence-based practices for UTI prevention. Practices such as adhering to clinical guidelines, keeping the perineal area clean, ensuring hand hygiene, improving management of urinary incontinence, and implementing hydration regimens have been shown to be effective in preventing UTIs (Montoya et al, 2016; Meddings et al., 2017; Wolff et al., 2017). Thus, this quality measure may also promote wider use of these practices.

Gaps in Performance among Specific Groups of Nursing Home Residents:

There is some evidence in the literature of empirical relationships between patient characteristics and UTI rates:

1. One study identified a positive relationship between female gender and UTI rate (Gucwa et al., 2016).
2. White race has been found to be a predictor of UTIs (Hefele et al., 2017; Castle et al., 2017).
3. Older age is associated with higher rates of UTIs (Castle et al., 2017).
4. Higher rates of Medicaid coverage in a facility are negatively associated with UTI rates among non-catheterized residents and positively associated with UTI rates among catheterized residents (Castle et al., 2017).

Overall, these findings are consistent with RTI's analysis, which found that individuals who were of non-Hispanic white race/ethnicity, older, and female were slightly more likely than their counterparts to have a UTI. In addition, our analyses also demonstrated that individuals who were not eligible for Medicaid were slightly more likely than individuals who were eligible for Medicaid to have a UTI.

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 (MDS) 3.0. In the pre-TEP survey, six out of 10 TEP members rated the UTI 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); has a demonstrated high-impact aspect of healthcare (e.g., affects large numbers); has external evidence of importance, such as consensus standards; and/or has evidence of disparities for the quality domain (RTI, 2019b).

The TEP noted that it was important to maintain public reporting of the UTI quality measure as it reflects a critical health outcome that warrants continued attention, citing significant negative outcomes of UTIs on nursing home residents' function, quality of life, and socialization. The TEP believed this quality measure was important for promoting quality improvement for the following reasons:

1. The UTI quality measure has an important role in promoting accountability and tracking urinary tract infection, allowing facilities to "identify patterns and implement solutions."

2. The presence of this quality measure keeps providers' efforts focused on improving accurate diagnosis and can also lead to improved antibiotic prescribing and antibiotic stewardship.

3. The TEP indicated the UTI quality measure encouraged investment in education and training for clinical and direct care staff.

Castle, N., Engberg, J. B., Wagner, L. M., & Handler, S. (2017). Resident and facility factors associated with the incidence of urinary tract infections identified in the nursing home Minimum Data Set. Journal of Applied Gerontology, 36(2), 173-194.

Cooper, D., McFarland, M., Petrilli, F., & Shells, C. (2019). Reducing inappropriate antibiotics for urinary tract infections in long-term care: A replication study. Journal of Nursing Care Quality, 34(1), 16-21.

Crnich, C. J. & Drinka, P. (2014). Improving the management of urinary tract Infections in nursing homes: It´s time to stop the tail from wagging the dog." Annals of Long Term Care: Clinical Care and Aging, 22(9), 32-36.

Gould CV, Umscheid CA, Agarwal R, Kuntz G, Pegues DA; the Healthcare Infection Control Practices Advisory Committee. Guideline for prevention of catheter-associated urinary tract infections 2009. Atlanta: Centers for Disease Control and Prevention, 2009. Available from

https://www.cdc.gov/infectioncontrol/pdf/guidelines/cauti-guidelines-H.pdf.

Herzig, C., Dick, A., Castle, N., & Stone, P. (2016). Policies and practices to reduce urinary tract infections in nursing homes [Abstract]. American Journal of Infection Control 44, S12.

Meddings, J., Saint, S., Krein, S. L., Gaies, E., Reichert, H., Hickner, A.,…Mody, L. (2017). Systematic review of interventions to reduce urinary tract infection in nursing home residents. Journal of Hospital Medicine, 12(5), 356

Montoya, A., Cassone, M., & Mody, L. (2016). Infections in nursing homes: Epidemiology and prevention programs. Clinics in Geriatric Medicine, 32(3), 585-607.

Nelson, S. T., & Flynn, L. (2015). Relationship between missed care and urinary tract infections in nursing homes. Geriatric Nursing, 36(2), 126-130.

RTI analysis of MDS 3.0 episode files for Quarter 1, 2011–Quarter 3, 2018 (programming reference: KH46\hf15_request_684_31_32.log)

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.

Smith, S. N., Greene, M. T., Mody, L., Banaszak-Holl, J., Petersen, L. D., & Meddings, J. (2018). Evaluation of the association between Nursing Home Survey on Patient Safety culture (NHSOPS) measures and catheter-associated urinary tract infections: Results of a national collaborative. BMJ Quality & Safety, 27(6), 464-473.

Trautner, B. W., Greene, M. T., Krein, S. L., Wald, H. L., Saint, S., Rolle, A. J.,...Mody, L. (2017). Infection prevention and antimicrobial stewardship knowledge for selected infections among nursing home personnel. Infection Control & Hospital Epidemiology, 38(1), 83-88.

Wolff, M. L., et al. (2016). An innovative quality assurance activity to reduce urinary tract infection rates in a green house skilled nursing setting." Annals of Long Term Care, 24(10), 17-20.

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, 1,096,778 long-stay residents in 14,520 nursing homes are included in the analysis. The national facility-level mean score for this measure in Quarter 3, 2018 was 2.8% and the median score was 1.9%, suggesting a slight positive skew. The interquartile range for this measure was 4.2%. 32.3% 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.

In Quarter 3, 2018:

k (facilities) 14,520

n (residents) 1,096,778

mean 2.8%

standard deviation (SD) 3.4%

min 0.0%

max 38.6%

Interquartile Range 4.2%

10th percentile 0.0%

20th percentile 0.0%

30th percentile 0.0%

40th percentile 1.3%

50th percentile 1.9%

60th percentile 2.6%

70th percentile 3.6%

80th percentile 4.9%

90th percentile 7.2%

SOURCE: RTI analysis of MDS 3.0 episode file for Quarter 3, 2018 (kh29_47\hf362_request_q3132_684.log).

Performance Over Time:

For comparison over time, we also present the distribution of facility-level scores on this quality measure in Quarter 2, 2013 (last presented to NQF during endorsement maintenance). During this time period, 1,108,999 long-stay residents in 13,640 nursing homes are included in the analysis. The national facility-level mean score for this measure in Quarter 2, 2013 was 6.2% and the median score was 5.3%, indicating a slight positive skew. The interquartile range for this measure was 6.4%. At least 10% of facilities had a perfect score of 0.0%. This analysis was restricted to facilities with at least 30 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 2, 2013:

k (facilities) 13,640

n (residents) 1,108,999

mean 6.2%

standard deviation (SD) 5.0%

min 0.0%

max 38.5%

Interquartile Range 6.4%

10th percentile 0.0%

20th percentile 2.0%

30th percentile 3.0%

40th percentile 4.2%

50th percentile 5.3%

60th percentile 6.5%

70th percentile 8.1%

80th percentile 9.9%

90th percentile 12.9%

SOURCE: RTI analysis of MDS 3.0 episode file for Quarter 2, 2013 (db277\db277_request.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. After an early increase in mean and median scores, the national facility-level mean and median scores have trended steadily downward since the adoption of the MDS 3.0, indicating a general improvement in performance over time. The mean score for this measure was 7.5% in Quarter 1, 2011, and the median score was 6.5%. In Quarter 3, 2018, the mean and median scores were 2.8% and 1.9%, respectively.

SOURCE: RTI analysis of MDS 3.0 episode files for Quarter 1, 2011–Quarter 3, 2018 (programming reference: KH46\hf15_request_684_31_32.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. Data are available and described in 1b2. The data is not an estimate based on samples; rather, it includes all 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 15,299 eligible facilities containing 1,104,673 residents eligible for inclusion in the measure (with both prior and target assessments); 14,520 facilities (95.0%) containing 1,096,778 residents (99.3%) had sufficient sample size (20 or more long-stay residents included in the denominator) to report on this measure.

We address the issue of disparities for this measure by examining whether there are differences in UTI rates for population groups that may reflect disparities in care, such as for population groups with social risk factors. We also examine facility-level performance gaps by concentration of racial/ethnic minority residents and Medicaid-eligible residents (may be a limited indicator of low socioeconomic status and/or indicative of the long-stay nursing home population receiving custodial care).

At the resident-level, we examined whether 4 social risk factors were associated with triggering the numerator for the UTI quality measure: 1) Medicaid eligibility; 2) race/ethnicity; 3) age (85 years and older); and 4) sex/gender. Details on how these variables were derived from the MDS 3.0 are provided in Section 2b3.3b. of the Testing Attachment.

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

1. We calculated the percentage of residents with and without each social risk factor who had a documented UTI per MDS 3.0 guidelines and the p-value for the Chi-squared test of differences among population groups;

2. We calculated the percentage of residents with a UTI individually among more granular race/ethnicity categories and conducted an Analysis of Variance (ANOVA) to test whether differences were statistically significant.

1. Differences in performance among specific groups of nursing home residents were small but statistically significant. Among residents who were eligible for Medicaid, 2.46% had a urinary tract infection and, among those ineligible for Medicaid, 3.35% had a urinary tract infection (?2(1) =324.39, p < .001). Among non-Hispanic white residents, 2.88% had a UTI, compared with 1.92% of non-white residents (?2(1) =765.46, p < .001). Among residents aged 85 years or older, 2.76% had a UTI, compared with 2.52% of younger residents (?2(1) = 60.50, p < .001). In addition, whereas 2.20% of male residents had a UTI, 2.84% of female residents had a UTI (?2(1) = 405.96, p < .001). Overall, individuals who were non-Hispanic white, older, female, and not eligible for Medicaid were slightly more likely than their counterparts to have a UTI. Additional information on social risk factor testing, including relevant citations that establish a conceptual framework from the literature review, is provided in the Testing Attachment.

2. At the resident-level, we also compared numerator triggering for this quality measure across more granular racial/ethnic groups. The highest percentage of long-stay residents with a urinary tract infection was found in American Indian or Alaska Native residents (3.0%), followed by White residents (2.9%), Black or African American residents (2.6%), Hispanic or Latino residents (2.0%), Native Hawaiian or Other Pacific Islander residents (1.7%), and Asian residents, who had the lowest percentage of long-stay residents with a UTI (1.6%). Using an ANOVA, differences in the proportion of residents with UTI by racial/ethnic group were found to be statistically significant (p<0.001) (RTI, 2019b).

At the facility-level, we compared facility performance on this quality measure in facilities with different proportions of non-Hispanic white and non-white residents. We examined differences in the percentage of long-stay nursing home residents with a UTI across two groups: facilities with proportions of non-Hispanic white residents that were greater than or equal to the median proportion (86.8%) among facilities with sufficient sample size to meet minimum public reporting requirements (>=20 episodes in the denominator), and facilities with a smaller proportion of non-Hispanic white residents (i.e., a larger proportion of non-white residents) than the median proportion. Facilities with a higher proportion of white residents had statistically significantly higher rates of UTI (3.3% compared to 2.4%; p<0.0001) (RTI, 2019c).

RTI also compared facility scores on this measure in facilities with different proportions of residents eligible for Medicaid. We examined differences in the percentage of long-stay nursing home residents with a urinary tract infection across two groups: facilities with a large proportion (greater than or equal to 75%) of residents eligible for Medicaid among facilities with sufficient sample size to meet the minimum public reporting threshold, and facilities with fewer than 75% of residents eligible for Medicaid. Facilities with a higher proportion of Medicaid-eligible residents had statistically significantly lower rates of UTI (2.7% compared to 3.7%; p<0.0001) (RTI, 2019c).

Our testing of social risk factors and their relationships to UTIs indicate that some factors (Medicaid eligibility, non-Hispanic white race/ethnicity, younger age, and male sex/gender were associated with lower UTI rates. Although associations with UTI were generally small, continued monitoring of potential disparities in UTI is critical to ensure continued utility of this measure to providers and consumers of nursing home care.

Performance Over Time:

For comparison over time, we also present facility scores on this measure in facilities with different proportions of non-White residents for Quarter 2, 2013 (last presented to NQF during endorsement maintenance), and found results similar to to 2018 data. Again, we examined differences in the percentage of long-stay nursing home residents with a UTI across two groups: facilities with proportions of White residents that were greater than or equal to the median proportion (87.0%) among facilities with sufficient sample size to meet minimum public reporting requirements (>=30 episodes in the denominator), and facilities with fewer White residents than the median. Facilities with a higher proportion of White residents had statistically significantly higher rates of urinary tract infection (6.9% compared to 5.0%; p<0.0001).

Additional analysis previously presented to NQF also examined facility measure performance in facilities stratified by the proportion of residents who are Medicaid eligible. Specifically, facilities were stratified into two groups: facilities with a large proportion (greater than or equal to 75%) of residents who are Medicaid eligible among facilities that meet the minimum denominator threshold for public reporting, and those that had less than 75 percent of residents who were Medicaid eligible. Facilities with a higher proportion of Medicaid-eligible residents had statistically significantly lower rates of UTI (6.0% compared to 7.1%; p<0.0001) (RTI, 2013).

1. RTI, 2019a. Results from Testing Attachment. RTI analysis of MDS 3.0 Data, Q3, 2018 (programming reference: KH42/hf11 _request_684_31_32.log, hf19_request_31_32.log, hf\hf20\hf20_request_31_32.log).

2. RTI, 2019b. RTI analysis of MDS 3.0 Data, Q3, 2018 (programming reference: kh29_47/hf365 _request_q3132_684.log)

3. RTI, 2019c. RTI analysis of MDS 3.0 Data, Q3, 2018 (programming reference: kh29_47/hf387 _request_q3132_684.log)

4. RTI, 2013. RTI analysis of MDS 3.0 Data, Q2, 2013 (programming reference: qm_quarter_11_12/complete/ae_request1_018.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. Data are available and described 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):*

Infectious Diseases (ID), Renal

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 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.

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 a urinary tract infection within the last 30 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 residents in the denominator sample with a selected target assessment that indicates urinary tract infection within the last 30 days (I2300 = [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) 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 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).

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)*

If the target assessment is an OBRA Admission Assessment, PPS 5-Day Assessment, or PPS Readmission/Return Assessment, the resident is excluded from the denominator for this quality measure. A resident is also

excluded if the target assessment indicates that data is missing for the data element assessing urinary tract infection in the last 30 days.

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.)*

A resident is excluded from the denominator if:

1. The target assessment is an OBRA Admission Assessment (A0310A = [01]) or a PPS 5-Day Assessment (A0310B = [01]) or a PPS Readmission/Return Assessment (A0310B = [06]).

2. The target assessment indicates that the value for the data element regarding urinary tract infection in the last 30 days is missing (I2300 = [-]).

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)

No risk adjustment or risk stratification

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.)*

Step 1: Identify the total number of long-stay residents with an episode during the quarter selected with a qualifying target assessment (OBRA, PPS, or discharge) and who do not meet the exclusion criteria (i.e., if the target assessment is an OBRA Admission Assessment, PPS 5-Day Assessment, or PPS Readmission/Return Assessment, or if I2300 = [-] on the target assessment).

Step 2: Starting with the set of residents identified in Step 1, determine the total number of long-stay residents with a selected target assessment (OBRA, PPS, or discharge) that meets the numerator inclusion criteria.

Step 3: Divide the results of step 2 by the results of step 1.

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

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

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 estimated based on samples; rather, the data include all 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).

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

NQF_MeasSubm_MeasTesting_Formv6.5_0684_UTI_01.17.2014final.docx,NQF_0684_UTI_Testing_Form_07242019_final.docx

2.1 For maintenance of endorsement

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

Yes

2.2 For maintenance of endorsement

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

Yes

2.3 For maintenance of endorsement

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

Yes - Updated information is included

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

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

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

Measure Number (*if previously endorsed*)**:** 0684
Measure Title: Percent of Residents with a Urinary Tract Infection (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 (Saliba & Buchanan, 2008).

2. RTI facility-level analyses of MDS 3.0 data sample included all facilities with sufficient sample size (n ≥ 20 residents) to publicly report this measure in Quarter 3, 2018 (k = 14,520), unless otherwise noted (RTI International, 2019).[1]

RTI International (2019). RTI analysis of MDS 3.0 data for Quarter 3 2018 (programming reference: kh29_47\hf354_request_q3132_684.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. 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.

[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.

2. The sample for the RTI analysis of MDS 3.0 includes all long-stay residents that meet the denominator inclusion criteria for this measure in facilities with sufficient sample size ($n \geq 20$, $k = 14{,}520$) to report this measure ($n = 1{,}096, 778$) in Quarter 3, 2018 (RTI International, 2019).

RTI International (2019). RTI analysis of MDS 3.0 data for Quarter 3 2018 (programming reference: kh29_47\hf354_request_q3132_684.log, kh29_47\hf354_request_q3132_682.log)

Table 1a below presents the characteristics of long-stay residents counted in the denominator for this measure in Quarter 3, 2018 *before* applying facility sample size restrictions and *without* excluding those residents who did not have both a prior and target assessment ($n = 1{,}104{,}673$); 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 residents that *may* be in the denominator regardless of changes in facility census. ***Table 1b*** offers the characteristics on the residents who are counted in the denominator *after* applying facility sample size restrictions in Quarter 3, 2018 ($n = 1{,}096{,}778$), 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 #0684 Percent of Residents with a Urinary Tract Infection (Long Stay) (Quarter 3, 2018)

Resident characteristics	Frequency (n)	Total Observations (N)	Percentage (%)
Sex			
Female	712,653	1,104,673	64.5%
Male	392,020	1,104,673	35.5%
Race/Ethnicity		1,104,673	
White Only	815,036	1,104,673	73.8%
Black or African American Only	171,331	1,104,673	15.5%
Hispanic or Latino Only	63,268	1,104,673	5.7%
Asian Only	21,954	1,104,673	2.0%
American Indian/Alaska Native Only	5,147	1,104,673	0.5%
Native Hawaiian or Other Pacific Islander Only	1,623	1,104,673	0.2%
Multi-race	3,355	1,104,673	0.3%
Medicare-Medicaid Dual Eligibility		1,104,673	
Dual-Eligible	798,534	1,104,673	72.3%
Non-Dual	226,602	1,104,673	20.5%
Missing	79,537	1,104,673	7.2%
Age			
<65	181,048	1,104,673	16.4%
65-74	202,469	1,104,673	18.3%
75-84	289,903	1,104,673	26.2%
85+	431,253	1,104,673	39.0%
Diagnoses			
Arthritis	328,847	1,100,591	29.9%
Osteoporosis	140,690	1,100,609	12.8%
Hip Fracture	18,330	1,004,739	1.8%
Other Fracture	32,255	1,004,720	3.2%

Resident characteristics	Frequency (n)	Total Observations (N)	Percentage (%)
Depression	524,829	1,004,637	52.2%
Stroke	126,897	1,004,696	12.6%
Alzheimer's Disease	159,464	1,004,705	15.9%
Non-Alzheimer's Dementia	490,486	1,004,654	48.8%
Malnutrition or at risk for malnutrition	55,081	1,104,631	5.0%
Cancer	71,233	1,100,574	6.5%
Anemia	298,292	1,004,582	29.7%
Heart Failure	207,639	1,004,700	20.7%
Hypertension	772,839	1,004,606	76.9%
Diabetes Mellitus	381,945	1,104,590	34.6%
Anxiety Disorder	351,495	1,104,527	31.8%
Asthma, Chronic Obstructive Pulmonary Disease, or Chronic Lung Disease	229,946	1,004,708	22.9%

Source: RTI analysis of Q3, 2018 MDS 3.0 data (programming reference: kh29_47\hf13_request_684_31_32.log)

Table 1b. Characteristics of Long-Stay Residents Included in Analyses of NQF #0684 Percent of Residents with a Urinary Tract Infection (Long Stay) (Quarter 3, 2018)

Resident characteristics	Frequency (n)	Total Observations (N)	Percentage (%)
Sex			
Female	707,343	1,096,778	64.5%
Male	389,435	1,096,778	35.5%
Race/Ethnicity			
White Only	808,833	1,096,778	73.7%
Black or African American Only	170,976	1,096,778	15.6%
Hispanic or Latino Only	63,710	1,096,778	5.8%
Asian Only	21,936	1,096,778	2.0%
American Indian/Alaska Native Only	5,484	1,096,778	0.5%
Native Hawaiian or Other Pacific Islander Only	2,194	1,096,778	0.2%
Multi-race	3,290	1,096,778	0.3%
Medicare-Medicaid Dual Eligibility			
Dual-Eligible	795,632	1,096,778	72.5%
Non-Dual	223,455	1,096,778	20.4%
Missing	77,691	1,096,778	7.1%
Age			
<65	180,174	1,096,778	16.4%
65-74	201,536	1,096,778	18.4%
75-84	288,060	1,096,778	26.3%
85+	427,008	1,096,778	38.9%

Resident characteristics	Frequency (n)	Total Observations (N)	Percentage (%)
Diagnoses			
Arthritis	326,574	1,092,828	29.9%
Osteoporosis	139,642	1,092,846	12.8%
Hip Fracture	18,089	998,068	1.8%
Other Fracture	31,884	998,050	3.2%
Depression	521,668	997,964	52.3%
Stroke	126,125	998,024	12.6%
Alzheimer's Disease	158,548	998,033	15.9%
Non-Alzheimer's Dementia	487,611	997,982	48.9%
Malnutrition or at risk for malnutrition	54,712	1,096,736	5.0%
Cancer	70,650	1,092,814	6.5%
Anemia	296,653	997,914	29.7%
Heart Failure	206,180	998,029	20.7%
Hypertension	767,986	997,934	77.0%
Diabetes Mellitus	379,808	1,096,697	34.6%
Anxiety Disorder	349,318	1,096,632	31.9%
Asthma, Chronic Obstructive Pulmonary Disease, or Chronic Lung Disease	228,523	998,035	22.9%

Source: RTI analysis of Q3, 2018 MDS 3.0 data (programming reference: kh29_47\hf13_request_684_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 urinary tract infection and available in the MDS 3.0, including race/ethnicity, Medicaid status, gender, and age. We selected these

resident-level social risk factors based on literature showing that UTI can vary by gender, race/ethnicity, Medicaid status, and age.

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 of the UTI item, the agreement between the MDS 2.0 item, coded by facility nurses, and the MDS 3.0 item, coded by gold-standard nurses was examined. Saliba and Buchanan (2008) present UTI rates using the MDS 2.0 and MDS 3.0 items at the resident- and facility-level, as well as Cohen's kappas, which 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. In their testing of the MDS 3.0, RAND calculated the UTI rate using the MDS 3.0 and the MDS 2.0, both at the individual resident-level and at the facility-level (Saliba & Buchanan, 2008). At the resident-level, the UTI rate using the MDS 2.0 was 10.0% and using the MDS 3.0 was 7.5%. At the facility-level, the MDS 2.0 rate of UTIs was 10.2% and the MDS 3.0 rate was 7.3%. Correlation between the MDS 2.0 and MDS 3.0 measures was strong at both the resident- (ρ = 0.71) and facility-level (ρ = 0.80). The Kappa for gold-standard to facility-nurse agreement on the MDS 3.0 and MDS 2.0 item was 0.70. Kappa is a statistical measure of inter-rater agreement for qualitative data, ranging from 0.0 to 1.0. A rating of 0.70 is considered "substantial agreement." These results are indicative of data element reliability.

Performance Measure Score Reliability

2.a. Signal-to-noise: The signal-to-noise ratio for this measure was 0.191 ($p < 0.001$) indicating that 19.1% 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 moderate (r = 0.42, ρ = 0.37, p < .001), and the ICC was 0.42 (p < .001) (RTI International, 2019b). Although approximately one-third of all facilities have values of 0% for this quality measure, this analysis provides moderate evidence of internal reliability because the variation in scores is sufficient: as shown in **Table 8** in **Section 2b4.2** below, the 50th percentile score is 1.9% and the 90th percentile score is 7.2%.

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). RTI analysis of Q3, 2018 MDS 3.0 data (programming reference: Kh29_47/hf363_request_q3132_684.log)

RTI International (2019b). RTI analysis of Q3, 2018 MDS 3.0 data (programming reference: kh29_47\hf14_request_684_31_32.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 show that this measure has moderate reliability. Using the measure of signal-to-noise ratio, the analysis shows about 19% of the variance in scores for this measure were explained by inter-facility characteristics, which is acceptable in the context of the variation in case mix in clinical settings, especially in nursing homes. This measure has moderate internal reliability as measured by the split-half correlation and Intraclass Correlation Coefficient (ICC) of 0.42, less than the threshold for strong reliability but acceptable.

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 #0684) was correlated with that facility's performance on the related quality measures 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 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 #0684 and its percentile rank on NQF #0686 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 #0684 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 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), which is used to determine numerator triggering for this measure may have 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. 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 #0684. TEP members discussed the current measure specifications, potential risk adjustment factors and the effectiveness of the measure in capturing quality of care, 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. Available at https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/MMS/TEP-Current-Panels.html.

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 #0684 (Percent of Residents with a Urinary Tract Infection (Long Stay)) and #0686 (Percent of Residents Who Have/Had a Catheter Inserted and Left in Their Bladder (Long Stay)) and found a positive but weak (ρ = 0.11, 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 #0684 (Percent of Residents with a Urinary Tract Infection (Long Stay)) and NQF #0685 (Percent of Low Risk Residents Who Lose Control of their Bowels or Bladder (Long Stay)) and found a positive (ρ=0.03) and statistically significant (p<0.001) relationship.

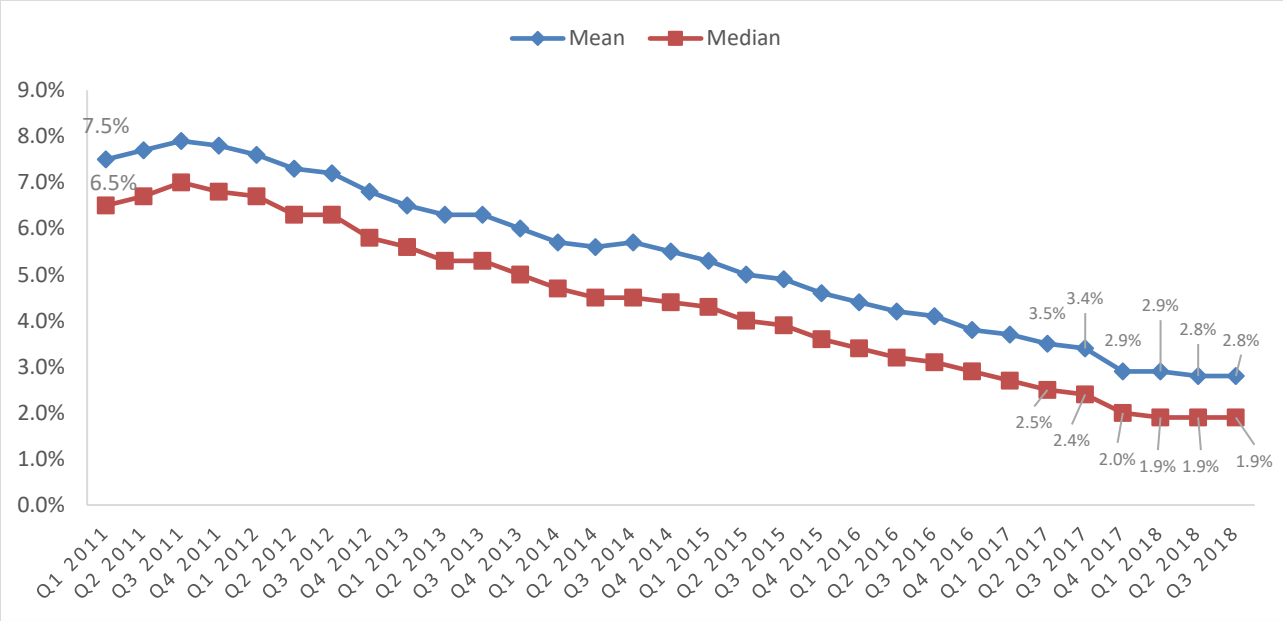
RTI International (2019a). RTI analysis of Q3, 2018 MDS 3.0 data (programming reference: KH33\hf359_request_q3132_684.log)

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 #0684. The proportion of variance in this measure explained by the state in which facilities are located is 1.4% ($p < 0.001$). The interquartile range of state-level scores is 4.2% (RTI International, 2019b).

RTI International (2019b). RTI analysis of Q3, 2018 MDS 3.0 data (programming reference: KH33\hf357_request_q3132_684.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, this also indicates that facilities may have improved practices related to genitourinary care, including prevention of urinary tract infections, during this period.

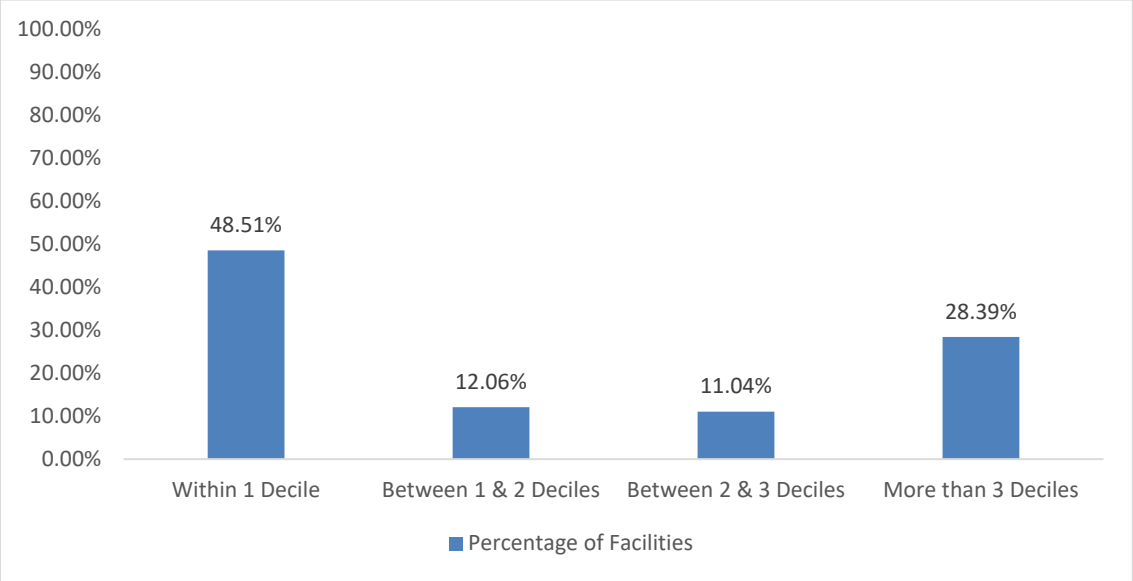
Figure 1. Seasonal (Quarterly) Variation, NQF #0684 Percent of Residents with a Urinary Tract Infection (Long Stay), Quarter 1, 2011 – Quarter 3, 2018



Source: RTI analysis of MDS 3.0 episode files for Quarter 1, 2011–Quarter 3, 2018 (programming reference: KH46\hf15_request_684_31_32.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. Most (48.5%) facilities are in the same decile in both quarters. Shifts of more than 3 deciles were less common, occurring for approximately 28.4% 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 #0684 Percent of Residents with a Urinary Tract Infection (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, 2018 and Q3, 2018 MDS 3.0 data (programming reference: KH32/hf358_request_q3132_684.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 many more facilities with quality measure scores that were statistically significantly ($p \leq .05$) lower than the national mean of 2.83% than those with scores that were statistically significantly higher than the national mean (5,229 versus 513), 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 decreases with the number of residents, except among the largest facilities (9^{th} and 10^{th}) deciles. Increases in the facility-level sample size lead to reductions in the standard error of facility-level scores, but larger facilities might have greater stabilities due to their larger sample size, which is less affected by a single infection. Changes in the reliability of this measure for the larger facilities may be accounted for by the greater statistical reliability that accompanies increased sample size as well as the increased stability.

Overall, 39.5% 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 providing evidence of validity for NQF #0684.

Table 2. Proportion of Facilities with Scores Significantly Different from the National Facility-Level Mean, Stratified by Facility Denominator Size for NQF #0684 Percent of Residents with a Urinary Tract Infection (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 33)	1,594	768 (48.2%)	41 (2.6%)	809 (50.8%)
2nd Decile (*n* = 34 to 41)	1,322	600 (45.5%)	20 (1.5%)	620 (46.9%)
3rd Decile (*n* = 42 to 50)	1,599	587 (36.7%)	76 (4.8%)	663 (41.5%)
4th Decile (*n* = 51 to 58)	1,371	476 (34.7%)	32 (2.3%)	508 (37.1%)
5th Decile (*n* = 59 to 67)	1,507	496 (32.9%)	53 (3.5%)	549 (36.4%)
6th Decile (*n* = 68 to 75)	1,319	386 (29.3%)	53 (4.0%)	439 (33.3)
7th Decile (*n* = 76 to 86)	1,553	448 (28.8%)	66 (4.2%)	514 (33.1%)
8th Decile (*n* = 87 to 98)	1,360	361 (26.5%)	57 (4.2%)	418 (30.7%)
9th Decile (*n* = 99 to 124)	1,457	549 (37.7%)	48 (3.3%)	597(41.0%)
10th Decile (*n* = 125 to 731)	1,438	558 (31.2%)	67 (4.7%)	625 (43.5%)
Total (*n* = 20 to 731)	**14,520**	**5,229 (36.0%)**	**513 (3.5%)**	**5,742 (39.5%)**

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: KH38\hf364_request_q3132_684.log)

1.f. Average change in performance across years: ***Table 3*** presents the changes in provider performance scores from year to year, fiscal year (FY) 2014 – FY 2018. On average, provider scores changed by less than 0.01 percentage points on NQF #0684. Few facilities experienced a change in performance by 0.05 percentage points or greater and over 90% of provider scores changed by 0.07 percentage points or fewer 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 0.007 percentage points, which is similar to previous mean facility score changes between other years when there was no change in clinical coding guidelines (0.006 – 0.009). Based on these findings, we include that there was no substantial change in provider score differences between years, including years when clinical coding guidelines did change (considering scores between FY 2017 and FY 2018, as the coding guideline changed at the beginning of FY 2018). 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 #0684.

Table 3. Distribution of Differences in Facility Performance Scores on NQF #0684 Percent of Residents with a Urinary Tract Infection (Long Stay) Across Years, FY 2014 – 2018

					Percentiles									
Difference	*K*	Mean	S.D.	Min	1st	5th	10th	25th	50th	75th	90th	95th	99th	Max
2014 – 2015	14,995	0.0058	0.035	-0.31	-0.09	-0.05	-0.03	-0.01	<0.01	0.02	0.05	0.06	0.11	0.33
2015 – 2016	14,987	0.0088	0.033	-0.69	-0.08	-0.04	-0.03	-0.01	0.01	0.03	0.05	0.07	0.11	0.57
2016 – 2017	15,012	0.0071	0.031	-0.32	-0.08	-0.04	-0.02	-0.01	0.01	0.02	0.04	0.06	0.10	0.46
2017 – 2018	14,969	0.0080	0.029	-0.52	-0.06	-0.03	-0.02	-0.01	0.01	0.02	0.04	0.06	0.10	0.42

Source: RTI analysis of MDS 3.0 Data, Q4, 2013 through Q3, 2018 (programming reference: hf\hf18\hf18_request_q_31_32_684.log)

1.g. Face validity: The majority of TEP members explicitly affirmed the face validity of NQF #0684. The TEP supported continued public reporting of the measure, as it allows providers to track their performance not only in correctly diagnosing UTIs, but also in antibiotic stewardship, which is closely linked to UTI management. Most TEP members agreed that the measure facilitated a declining trend in UTI rates

over time, and reflected quality of care in nursing homes. Some TEP members suggested looking at the relationship between UTI and function, hospice care, and dementia to see if they might be appropriate risk adjustors (see **Section 2b3.3a** for analysis of candidate factors); however, TEP members voiced support for the face validity of NQF #0684 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. Available at https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/MMS/TEP-Current-Panels.html.

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 urinary tract infections. The testing results indicated high validity according to analysis of seasonal variation. Facility-level measure scores do not vary substantially from quarter to quarter corresponding to changes in seasons; thus, seasonality is not a threat to validity for this measure. The testing results also indicate high validity according to analysis of change in measure performance over years, confidence interval analysis, and variation by state (with a low proportion of variance explained by state). The measure also showed moderate validity according to correlations with related quality measures; i.e., facilities' scores on this QM are positively correlated with their scores on #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)), providing some evidence supporting convergent validity. The testing did suggest some instability in facility decile ranking over time, potentially attributable to lower prevalence and narrow distribution. However, change in measure performance over years was small, even after the revision of item coding guidelines, indicating that changes in facility scores are not due to change in coding practice. 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) There are missing data in the response to urinary tract infection item in the target assessment.

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

Exclusion criterion 2 relating to missing data is assessed in greater detail in **Section 2b6**.

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 13,352 residents (1.19% of 1,118,025 long-stay residents in Quarter 3, 2018) were excluded from this quality measure based on the measure denominator exclusions as described above. Please note that exclusion criteria are not mutually exclusive.
- 12,911 residents (1.15% of long-stay residents in Quarter 3, 2018) were excluded because their target assessments were either admission, PPS 5-day, or readmission/return assessments. 443 residents (0.04% of long-stay residents in Quarter 3, 2018) were excluded because data was missing on the item I2300 on the MDS.

Source: RTI analysis of Q3, 2018 MDS 3.0 data (programming reference: kh29_47\hf360_request_q3132_684.log)

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 the type of assessment that serves as the resident's target assessment. Excluding these assessments to capture urinary tract infection information is appropriate because these assessments occur immediately following time the resident has spent not under the facility's care, and therefore would not accurately reflect the quality of the facility's care.

A small number of exclusions occur because of missing data. The impact of missing data on this quality measure is presented in detail in **Section 2b6**.

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** Click here to enter number of factors **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.

Not applicable. This measure is not risk-adjusted.

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.

The measure is not risk adjusted through a statistical model nor through stratification.

Urinary tract infections have relatively high prevalence across the continuum of care and there are no obvious conditions for which risk adjustment is appropriate. During the development of the MDS 3.0 measure, no major conditions were identified that were appropriate for risk adjustment and clearly associated with UTI. Urinary tract infections are often associated with catheter use, which is often inappropriate (Gould et al., 2009). Thus, risk adjusting for catheter use would not be desirable. The 2019 TEP expressed concern that mobility limitations would be a risk factor for UTIs that could warrant some

type of risk adjustment, which we explored further (see below) before deciding not to perform testing on potential clinical risk factors or risk-adjustment specifications. Discussion of the rationale for risk-adjustment testing for social risk factors is presented in **2b3.3a**.

Gould CV, Umscheid CA, Agarwal R, Kuntz G, Pegues DA; the Healthcare Infection Control Practices Advisory Committee. Guideline for prevention of catheter-associated urinary tract infections 2009. Atlanta: Centers for Disease Control and Prevention, 2009. Available from http://www.cdc.gov/ncidod/dhqp/pdf/guidelines/CAUTI_Guideline2009final.pdf.

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?

Risk Adjustor Selection – Conceptual Rationale and Statistical Testing

Clinical Risk Factors

At the suggestion of the 2019 TEP, we conducted testing on several clinical factors, including hospice care and functional status. Per the Long-Term Care Facility Resident Assessment Instrument 3.0 User's Manual, the hospice item on the MDS (O0100K2) identifies residents who were in a hospice program for "…terminally ill persons where an array of services is provided for the palliation and management of terminal illness and related conditions" within the last 14 days (Centers for Medicare & Medicaid Services, 2018). Tested function items from Section G of the MDS pertain to assistance with self-performance on Activities of Daily Living (ADLs), including:

- Bed mobility (G0110A1) – how resident moves to and from lying position, turns side to side, and positions body while in bed or alternate sleep furniture.
- Transfer (G0110B1) – how resident moves between surfaces including to or from: bed, chair, wheelchair, standing position (excludes to/from bath/toilet).
- Walk in room (G0110C1) – how resident walks between locations in his/her room.
- Walk in corridor (G0110D1) – how resident walks in corridor on unit.
- Toilet use (G0110I1) - how resident uses the toilet room, commode, bedpan, or urinal; transfers on/off toilet; cleanses self after elimination; changes pad; manages ostomy or catheter; and adjusts clothes. Does not include emptying of bedpan, urinal, bedside commode, catheter bag or ostomy bag.

We assessed the relationship between the facility mean proportion of residents receiving hospice care and performance on the UTI measure across QM score deciles. Nationally, the facility mean proportion of residents receiving hospice care was 8.4% (standard deviation [SD] = 7.2%). As QM score increases, the mean proportion of residents on hospice increases slightly from 8.5% to 9.5%. However, the mean proportion does not increase monotonically across QM score deciles, and the facility mean proportion of residents receiving hospice care ranged from 6.6% to 9.5% overall.

We also assessed the relationship between the facility mean proportion of residents who were totally dependent (i.e., for activities performed 3 or more times in the past 7 days, nursing home staff performed the activity each time) on each of the function items described above and performance on the UTI measure across QM score deciles.

- Bed mobility
    - Facility mean proportion of residents who were totally dependent was 8.6% (SD = 11.2%). As QM score increases, the mean proportion of residents who were totally dependent increases from 8.5% to 9.5%. However, the mean proportion does not increase monotonically across QM score deciles, and the facility mean proportion of residents who were totally dependent ranged from 8.4% to 9.5% overall.
- Transfer

    - Facility mean proportion of residents who were totally dependent was 20.0% (SD = 13.7%). As QM score increases, there is a small increase in the mean proportion of residents who were totally dependent, from 19.2% to 20.9%. However, the mean proportion does not increase monotonically across QM score deciles, and the facility mean proportion of residents who were totally dependent ranged from 19.2% to 21.0% overall.
- Walk in room
    - Facility mean proportion of residents who were totally dependent was 57.3% (SD = 16.4%). As QM score increases, the mean proportion of residents who were totally dependent increases from 56.1% to 58.2%. However, the mean proportion does not increase monotonically across QM score deciles, and the facility mean proportion of residents who were totally dependent ranged from 56.1% to 58.6% overall.
- Walk in corridor
    - Facility mean proportion of residents who were totally dependent was 61.2% (SD = 16.2%). As QM score increases, there is a small increase in the mean proportion of residents who were totally dependent, from 59.9% to 63.1%. However, the mean proportion does not increase monotonically across QM score deciles.
- Toilet use
    - Facility mean proportion of residents who were totally dependent was 16.8% (SD = 16.6%). As QM score increases, the mean proportion of residents who were totally dependent increases from 16.1% to 18.4%. However, the mean proportion does not increase monotonically across QM score deciles.

Results of testing demonstrated that there were not strong relationships between performance on the UTI quality measure and proportion of residents receiving hospice care or who were totally dependent on the functional items. Therefore, we did not proceed with risk adjustment for clinical factors and did no further testing for risk adjustment (RTI International, 2019).

Centers for Medicare & Medicaid Services (2018, October). Long-Term Care Facility Resident Assessment Instrument 3.0 User's Manual, Version 1.16. https://downloads.cms.gov/files/1-MDS-30-RAI-Manual-v1-16-October-1-2018.pdf

Source: RTI analysis of Q3, 2018 MDS 3.0 data (programming reference: ljc\ljc73_request_684_31_32.log)

Social Risk Factors

We conducted a recent literature review on UTIs in nursing homes to determine whether other researchers had posited a conceptual basis for why social risk factors might influence the incidence of UTIs in nursing homes, such that the risk factor could not be addressed through nursing home care delivery (NQF, 2017). Some studies found an *empirical* association between social risk factors that could be measured by items available in the MDS 3.0 and UTI, but did not offer a *conceptual* basis for understanding how the inherent characteristics of the social risk factor (race/ethnicity, age, gender, and Medicaid coverage) would affect the development or avoidance of UTI. 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 the absence of any other risk adjustment based on clinical or social risk factors.

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. Available at

https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/MMS/TEP-Current-Panels.html.

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 age, Medicaid coverage, and race/ethnicity are influencing the incidence of UTIs in nursing homes. In the case of gender, the association between gender and UTIs 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 UTI:

- One study identified a positive association between female gender and the rate of UTIs (Gucwa et al. 2016).
- One study identified a link between both age and having Medicaid coverage with UTIs, with older residents having a higher risk of infection and higher Medicaid coverage in a facility negatively associated with UTI incidence (Castle et al., 2017).
- White race was also identified as a predictor of UTI 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 urinary tract infection, 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 of these differences were statistically significant, the differences across subpopulations are mostly small. Among residents who were eligible for Medicaid, 2.46% had urinary tract infections and,

among those ineligible for Medicaid, 3.35% had urinary tract infections ($\chi^2(1) = 324.39$, $p < .001$). For non-Hispanic white residents, 2.88% had urinary tract infections, compared with 1.92% of non-white residents ($\chi^2(1) = 765.46$, $p < .001$). For residents aged 85 years or older, 2.76% had urinary tract infections, compared with 2.52% of younger residents ($\chi^2(1) = 60.50$, $p < .001$). In addition, whereas 2.20% of the male residents had urinary tract infections, 2.84% of the female residents had urinary tract infections ($\chi^2(1) = 405.96$, $p < .001$).

Table 4. NQF #0684 Percent of Residents with a Urinary Tract Infection (Long Stay) by Social Risk Factors (Quarter 3, 2018)

Resident characteristic	Frequency of residents who had urinary tract infections (n)	Percentage of residents who had urinary tract infections (%)	Pearson chi^2 P-value
Race/Ethnicity			
Non-Hispanic white	23,499	2.88%	<0.001
Non-white	5,561	1.92%	
Age			
≥ 85	11,814	2.76%	<0.001
< 85	16,807	2.52%	
Gender			
Male	8,538	2.20%	<0.001
Female	20,083	2.84%	
Medicaid			
Medicaid	21,856	2.46%	<0.001
Non-Medicaid	3,922	3.35%	

Source: RTI analysis of MDS 3.0 Data, Q3, 2018 (programming reference:KH42/hf11 _request_684_31_32.log, hf19_request_31_32.log)

Overall, individuals who identified as non-white race/ethnicity, were older, female, and non-Medicaid eligible, were slightly more likely than their counterparts to have a urinary tract infection. For females, this result was expected, and the other results were also supported by the literature.

Given CMS's guidance to avoid having risk adjusters mask disparities in care associated with these factors, and instead consider 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 UTI measure, results indicated that, of the facilities with publicly reportable scores (≥20 residents in the denominator) for the current specification, approximately 75.1%, 49.3%, and 92.8% 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 UTI 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 #0684 Percent of Residents with a Urinary Tract Infection (Long Stay) (Quarter 3, 2018)

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	3,803	24.9%

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
Gender	7,742	50.7%
Medicaid status	1,100	7.2%

Source: RTI analysis of Q3, 2018 MDS 3.0 Data (programming reference: hf\hf20\hf20_request_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.

CMS (2018, August). Blueprint for the CMS Measures Management System. https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/MMS/Downloads/BlueprintVer14.pdf

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?

This measure, the long-stay urinary tract infection measure (NQF #0684), is not currently risk-adjusted. NQF #0684 Percent of Residents With a Urinary Tract Infection (Long Stay) was endorsed by NQF without denominator exclusion and model-based risk adjustment. During the development of the MDS 3.0 measure, no major conditions were identified that were appropriate for risk adjustment and clearly associated with UTI.

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

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 UTIs, RTI examined whether there is variation in the social risk factor among nursing homes (there is; see ***Table 6***), and the potential improvement in the risk model if social risk factors are included.

Table 6. Distribution of Percentage of Residents with Select Resident Characteristics Across Facilities (Quarter 3, 2018)

Resident characteristics	Facilities (*k*)	Mean % of residents	Std dev.	10th percentile	25th percentile	50th percentile	75th percentile	90th percentile
Age ≥ 85	14,496	40.7%	19.7%	15.1%	26.3%	40.0%	54.7%	67.4%

Source: RTI analysis of MDS 3.0 Data, Q3 2018 (programming reference: KH42/hf11_request_684_30_31.log)

As the analysis in ***Table 4*** above indicates, age of 85 years and older was also associated with urinary tract infections and the difference in the proportion of residents who trigger the numerator by age group is small but 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 a model risk adjusted for oldest old.

In the candidate model, the odds ratio for oldest old is 1.10 (95% CI = [1.07, 1.12]), and is statistically significant at the 0.05 level. Consistent with previous studies, the odds of urinary tract infections are 10% higher among residents aged 85 and older, compared to younger residents.

Additional information regarding goodness of fit for the logistic regression model is presented in ***Table 7***.

Table 7. Assessment of Model Performance for Alternate Risk Adjustment Specifications, NQF #0684 (Quarter 3, 2018)

Model Covariates	Hosmer-Lemeshow Chi^2, *P*-value	AIC	BIC	Pseudo R^2	Log Likelihood
Candidate model: age ≥ 85	--	260,183.3	260,207.1	0.0002	-130,089.6

NOTES: "-- " indicates that the Hosmer-Lemeshow test only has two distinct quantiles, and the significance of the test statistic cannot be computed or is misleading, based on other model selection statistics.

AIC: Akaike Information Criterion

BIC: Bayesian Information Criterion

Candidate model: Risk adjusted for oldest old

Source: RTI analysis of MDS 3.0 Data, Q32018 (programming reference: KH43/hf12_request_ 684_30_31.log)

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), pseudo R^2, and log-likelihood to examine the statistical risk model calibration. We also used the Hosmer-Lemeshow test for goodness of fit; however, when only 1 predictor is in the risk adjustment model, the Hosmer-Lemeshow test only has two quantiles, and the significance of the test statistic cannot be computed.

We then examined the model fit and calibration for the model presented in ***Table 7***. When we risk adjust for the oldest old, the c-statistic is 0.51. This indicates weak model performance.

Due to the results of our analyses, which are described above, we do not recommend risk adjusting this measure for social risk factors. There is almost no practical improvement in including risk adjustors in the model, and the added complexity would make the measure more difficult to interpret for the public.

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

Not applicable. This measure is not risk-adjusted.

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

Not applicable. This measure is not risk-adjusted.

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

Not applicable. This measure is not risk-adjusted.

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

Not applicable. This measure is not risk-adjusted.

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 risk adjustment by stratification on gender, race/ethnicity, and Medicaid eligibility is infeasible, that clinical factors lacked empirical support for inclusion in risk adjustment models, and that the risk adjusted model that accounts for older age has weak model performance and does not have sufficient predictive ability. The c-statistic for this model is 0.51, which is below the threshold for acceptability (i.e., >0.60). Therefore, we conclude that additional risk adjustment for social or clinical risk factors offers little practical improvement to the quality measure, and thus do not intend to add a risk adjustment strategy.

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 #0684, 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 8 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 2.8% in Quarter 3, 2018 with a median score of 1.9%. The interquartile range for this measure was 4.2%. Among facilities who were eligible to publicly report this measure, 32.3% (k = 14,520) had perfect scores of 0%; as shown in ***Figure 1*** in **Section 2b1**, the national mean facility-level score on this measure has decreased over time, suggesting that facilities may

have improved in practices related to genitourinary care, including prevention of urinary tract infections, after this measure was publicly reported.

Table 8. National Facility-Level Score Distribution, NQF #0684 Percent of Residents with a Urinary Tract Infection (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
14,520	2.8%	3.4%	0.0%	0.0%	1.9%	4.2%	7.2%	32.3%	4.2%

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: KH30/hf356_request_q3132_684.log; KH39/hf365_request_q3132_684.log)

Table 2 (**Section 2b1.3**) 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 (39.5%) 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. Increases in the facility-level sample size lead to reductions in the standard error of facility-level scores, thus, it appears that changes in the reliability of this measure for larger facilities are due to the greater statistical reliability that accompanies increased sample size.

Source: RTI analysis of Q3, 2018 MDS 3.0 data (programming reference: KH38\hf364_request_q3132_684.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 urinary tract infection measure, both overall and stratified by quality measure score quartile.

2. We analyzed whether missing data on the urinary tract item varied systematically by several resident-level characteristics which are associated with urinary tract infection, including gender, age, race/ethnicity, and Medicaid eligibility.

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)

Among all long-stay residents in Quarter 3, 2018 (*n*= 1,118,025), 0.04% (*n*=443) had missing data for at least one of the items necessary to calculate the UTI measure.

In addition, RTI examined the relationship between missing data for items used to calculate this measure by quality measure score quartile. ***Table 9*** shows the mean facility-level missing rate for items used to calculate this measure is lowest in the second score quartile, and highest in the worst score quartile, and ranges from 0.01%-0.08%. There is also a significant but weak correlation between missing data and quality measure scores ($r = 0.042$, $p < 0.001$).

Table 9. Distribution of Facility-Level Missing Rate by Measure Score Quartile, NQF #0684 Percent of Residents with a Urinary Tract Infection (Long Stay) (Quarter 3, 2018)

Quality Measure Score Quartile	Facilities (k)	Mean	Std dev.	10th	25th	50th	75th	90th
0–25 (Best)†	4,690	0.03%	0.68%	0	0	0	0	0
26–50	2,622	0.01%	0.20%	0	0	0	0	0
51–75	3,614	0.05%	0.65%	0	0	0	0	0
76–100 (Worst)†	3,594	0.08%	0.92 %	0	0	0	0	0
Total	14,520	0.05%	0.69%	0.0	0.0	0.0	0.0	0.0

Source: RTI analysis of Q3, 2018 MDS 3.0 data (programming referencekh29_47\hf361_request_q3132_684.log)

Our analysis showed that the average missing data rate across facilities was 0.054% (shown in the "Total" row of ***Table 9***) on items used to construct this measure. Fewer than 10% of facilities had any missing data at all on items used to construct this measure.

This analysis addresses the potential concern that missing data in the items used to construct this QM would lead to under-reporting urinary tract infection, resulting in lower (better) scores on this measure. Our analysis demonstrated that this does not appear to be an issue for this measure. The positive relationship between missing rate and percentage of residents who have a urinary tract infection suggests that facilities with higher rates of urinary tract infections may not do well in providing care related to genitourinary conditions or in preventing UTIs.

Table 10 summarizes the results of RTI's analysis of whether missing data varied across selected resident characteristics related to urinary tract infection. Specifically, we analyzed whether missing data varied systematically on the following characteristics: race/ethnicity, age greater than or equal to 85, gender, and Medicaid eligibility, and used Chi-Squared tests to determine whether these differences were statistically significant.

Table 10. Frequency of Missing Data by Select Resident Characteristics Among Long-Stay Residents, NQF #0684 Percent of Residents with a Urinary Tract Infection (Long Stay) (Quarter 3, 2018)

Risk Factor	Percent of Residents (%)	Frequency of Residents (n)	Pearson chi^2 P-value
Race/Ethnicity			
Non-Hispanic White	0.04%	329	0.258
Non-White/Multi-racial	0.04%	47	

Age			
Age ≥ 85	0.04%	254	0.268
Age < 85	0.04%	181	
Gender			
Male	0.04%	139	0.121
Female	0.04%	296	
Medicaid Eligibility			
Medicaid	0.04%	329	0.611
Non-Medicaid	0.04%	47	

**p<0.05*

Source: RTI analysis of Q3, 2018 MDS 3.0 data (programming reference kh29_47\hf16_request_684_31_32.log)

Table 10 shows that differences in missing data by race/ethnicity, age, gender, and Medicaid eligibility, were not statistically significant (p>0.05). Across all individual characteristics, rates of missing data were nearly identical among individuals, with no difference in percentage of missing data.

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 overall missing data rate for this measure was quite low (0.054%), and at least 90% of facilities were not missing data on the urinary tract item. Rates of missing data on the item used to construct this QM are very similar among individuals with selected characteristics related to urinary tract infection. 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 facilities; no issues are anticipated.

CMS provides coding directions for the UTI data element in the MDS 3.0 via the RAI Manual and other mediums, such as this YouTube video (https://www.youtube.com/watch?v=sZLjJMntcPQ) explaining the MDS 3.0 coding of Section I.

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)*.

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 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

? Geographic area and number and percentages of accountable entities and patients included: All United States Medicare/Medicaid certified nursing homes with eligible long-stay residents. In Quarter 3, 2018, there were 15,299 eligible facilities containing 1,104,673 residents eligible for inclusion in the measure (with both prior and target assessments); 14,520 facilities (95.0%) containing 1,096,778 residents (99.3%) had sufficient sample size (20 or more long-stay residents included in the denominator) to report on this measure.

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 patients included: All United States Medicare/Medicaid certified nursing homes with eligible long-stay residents. In Quarter 3, 2018, there were 15,299 eligible facilities containing 1,104,673 residents eligible for inclusion in the measure.

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 patients included: All United States Medicare/Medicaid certified nursing homes with eligible long-stay residents. In Quarter 3, 2018, there were 15,299 eligible facilities containing 1,104,673 residents eligible for inclusion in the measure.

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 #0684, Percent of Residents With a Urinary Tract Infection (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 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 Quality Measure (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. 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 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, and their Five Star Ratings.

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 QM reports and publicly reported information, are provided in the CASPER Reporting MDS Provider Users Guide, Section 11.

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 Five 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) Percent of Residents with a Urinary Tract Infection (Long Stay) (NQF #0684); and (2) Percent 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 used to inform discussion topics for the TEP meeting held on May 23, 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. Available at https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/MMS/TEP-Current-Panels.html.

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 this quality measure 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 it is a critical health outcome and that this quality measure promotes accurate diagnosis and tracking of UTIs as well as the appropriate use of antibiotics.

A majority of TEP members also noted that they use this QM to track facility performance and address concerns, including reviewing residents that trigger the QM, ensuring proper documentation, evaluating staff understanding of instructions in the Long-Term Care Facility Resident Assessment Instrument 3.0 User's Manual, addressing clinical practices, and comparing their facility to state and national benchmarks. In particular, with respect to the consumer perspective, one TEP member stated that this QM "provide[s] good information for follow up questions to [the] facility."

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 recent (October 2017) changes in UTI coding guidelines in the Long-Term Care Facility Resident Assessment Instrument 3.0 User's Manual. One TEP member suggested that mobility limitations be considered as a potential risk adjuster because "The increased pooling of urine/urinary stasis, the higher level at which the urethra sits in the prone patient, and the weaker extruder muscles that can lead to incomplete emptying all may increase risk of UTI in the immobile patient." Based on this suggestion, RTI conducted testing of several

MDS data elements assessing function. Results (presented in 2b3.3a. in the Testing Attachment) demonstrated that there were not strong relationships between measure performance and functional data elements. Therefore, we did not proceed with further risk adjustment testing.

Most TEP members (8 out of 10) reported no evidence or rationale for modifying the measure specifications to include exclusion criteria or to use a statistical risk model or stratification to adjust for resident social or clinical risk factors. TEP members noted that risk adjustment could mask important disparities, rather than help providers and consumers better assess facility performance.

A minority of TEP members questioned whether the UTI measure reflected quality of care in nursing homes and whether it could be used for improvement. They were concerned that nursing home staff may not apply evidence-based criteria to diagnose UTIs as intended and that some residents may develop UTIs despite a nursing home's best efforts to prevent them. However, other TEP members countered this argument, noting that the quality measure "allows for better accountability on the part of those diagnosing UTIs" (RTI, 2019).

National Quality Forum. (2019). Measure evaluation criteria. Retrieved from http://www.qualityforum.org/Measuring_Performance/Submitting_Standards.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. Available at https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/MMS/TEP-Current-Panels.html.

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 reviewed the measure favorably with respect to importance, usability and use, and validity. Additional empirical testing (provided in the Testing Attachment) conducted based on TEP feedback did not provide strong evidence to support measure modification. Based on our synthesis of the literature, our empirical testing, and TEP feedback, we contend 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. 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):

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. After an early increase in mean and median scores, the national facility-level mean and median scores have trended steadily downward since the implementation of the MDS 3.0 measure, indicating a general improvement in performance over time. The

mean score for this measure was 7.5% in Quarter 1, 2011, and the median score was 6.5%. In Quarter 3, 2018, the mean and median scores were 2.8% and 1.9%, respectively.

Geographic area and number and percentages of accountable entities and patients included:

All United States Medicare/Medicaid certified nursing homes with eligible long-stay residents. In Quarter 3, 2018, there were 15,299 eligible facilities containing 1,104,673 residents eligible for inclusion in the measure (with both prior and target assessments); 14,520 facilities (95.0%) containing 1,096,778 residents (99.3%) had sufficient sample size (20 or more long-stay residents included in the denominator) to report on this measure.

SOURCE: RTI analysis of MDS 3.0 episode files for Quarter 1, 2011–Quarter 3, 2018 (programming reference: KH46\hf15_request_684_31_32.log)

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.

The implementation of the MDS 3.0 version of the long-stay UTI measure demonstrated an improvement over the MDS 2.0 version of the same measure because it addressed potential causes of error by requiring more reliable evidence of UTI (CMS, 2009; Abt, 2007; Saliba and Buchanan, 2008; Stevenson, Moore, and Sleeper, 2007). Currently, this measure is the only one in the existing measure set that addresses the critical issue of infections in nursing homes, helping to drive quality improvements in the resident safety domain.

Of note, as this QM encourages staff education and supports accurate clinical diagnosis, this can also promote accurate treatment of UTIs. For example, there are negative implications associated with the "potential overuse of antibiotics to treat asymptomatic bacteriuria" and, during a May 2019 TEP, TEP members agreed that this QM promotes appropriate antibiotic use by improving antibiotic prescribing and antibiotic stewardship (RTI, 2019).

Abt Associates, Inc.; Stepwise Systems, Inc.; Qualidigm. Data Assessment and Verification (DAVE 2) project—MDS two-stage discrepancy findings, April–December 2006. Cambridge, MA: Abt Associates, Inc, 2007.

Centers for Medicare & Medicaid Services. Long-term care facility resident assessment instrument user's manual. Baltimore, MD. 2009. Available at http://www.cms.hhs.gov/NursingHomeQualityInits/25_NHQIMDS30.asp#TopOfPage.

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.

Saliba D, Buchanan J. 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, Apr 2008. Available from http://www.cms.hhs.gov/NursingHomeQualityInits/Downloads/MDS30FinalReport.pdf.

Stevenson KB, Moore JW, Sleeper B. Validity of the Minimum Data Set in identifying urinary tract infections in residents of long-term care facilities. J Am Geriatr Soc. 2007;52: 707-711.

4b2.2. Please explain any unexpected benefits from implementation of this measure.

This is not applicable; there are no unexpected benefits. Positive unexpected findings, including improved antibiotic prescribing and antibiotic stewardship, are described in 4b2.1.

5. Comparison to Related or Competing Measures

If a measure meets the above criteria and there are endorsed or new related measures (either the same measure focus or the same target population) or competing measures (both the same measure focus and the same target population), the measures are compared to address harmonization and/or selection of the best measure.

5. Relation to Other NQF-endorsed Measures

Are there related measures (conceptually, either same measure focus or target population) or competing measures (conceptually both the same measure focus and same target population)? If yes, list the NQF # and title of all related and/or competing measures.

Yes

5.1a. List of related or competing measures (selected from NQF-endorsed measures)

0138 : National Healthcare Safety Network (NHSN) Catheter-associated Urinary Tract Infection (CAUTI) Outcome Measure

0281 : Urinary Tract Infection Admission Rate (PQI 12)

5.1b. If related or competing measures are not NQF endorsed please indicate measure title and steward.

Measures with Similar Focus and Similar Target Population that are Not NQF-Endorsed

- Improvement in Urinary Tract Infection (This is a measure listed on the CMS Measures Under Development list for Home Health Quality Reporting)(1)
- Development of urinary tract infection (This is a measure listed on the CMS Measures Under Development list for Home Health Quality Reporting) (1)

Measures with Similar Focus and Different Target Population that are Not NQF-Endorsed

- Urinary tract infection (UTI) admission ratea (area-level): rate per 100,000 population (AHRQ, not endorsed)

(1) Source: http://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/QualityMeasures/CMS-Measures-Inventory.html

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.

0138 : National Healthcare Safety Network (NHSN) Catheter-associated Urinary Tract Infection (CAUTI) Outcome Measure This measure provides the Standardized Infection Ratio (SIR) of healthcare-associated, catheter-associated urinary tract infections (UTI) among patients in bedded inpatient care locations, except level II or level III neonatal intensive care units (NICU). This includes acute care general hospitals, long-term acute care hospitals, rehabilitation hospitals, oncology hospitals, and behavior health hospitals. The SIR is the ratio of the total number of observed healthcare-associated CAUTIs among patients in bedded inpatient care locations (excluding patients in Level II or III neonatal ICUs) to the total number of predicted healthcare-associated CAUTI among inpatient care locations under surveillance for CAUTI during the data period, based on the national CAUTI baseline. CAUTI prevention is important as it is associated with increased morbidity and mortality and higher healthcare costs. Although related to UTIs, CAUTIs reflect a distinct issue that may require

different clinical intervention; as such, providers' efforts to prevent UTIs and CAUTIs may vary. For example, CAUTI prevention includes reducing the number of unnecessary indwelling catheters inserted, removing indwelling catheters at the earliest possible time, securing catheters to the patient's leg to avoid bladder and urethral trauma, keeping the urine collection bag below the level of the bladder, and utilizing aseptic technique for urinary catheter insertion. Nursing home factors and best practices associated with UTI prevention are described in Section 1b.1. above. In addition, it may be challenging to measure CAUTI in nursing homes due to concerns about the availability of onsite laboratory testing; subsequently, reportability of a nursing home CAUTI measure may be substantially diminished. 0281 : Urinary Tract Infection Admission Rate (PQI 12) This measure reports the rate of admissions with a principal diagnosis of urinary tract infection per 100,000 population, ages 18 years and older. Patients with kidney or urinary tract disorder admissions, other indications of immunocompromised state admissions, obstetric admissions, and transfers from other institutions are excluded from the measure. Presence of a urinary tract infection is based on a principal diagnosis code (ICD-9) for UTI. UTIs in the adult population may generally be treated in ambulatory/outpatient care settings. However, when treatment is inadequate or delayed, patients may develop more severe clinical infections; as a result, they may be more likely to present at an emergency department and, subsequently, require inpatient admission. Therefore, access to sufficient outpatient care may be key to reducing urinary tract infection admissions. Although NQF #0281 and Percent of Residents With a Urinary Tract Infection (Long Stay) (NQF #0684) both capture UTI rates, they are intended for use in disparate populations (adults utilizing inpatient acute care vs. long-stay nursing home residents).

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.)

This is not applicable. There are no competing measures for this QM. None of the measures listed in the response to Question 5 above have the same measure focus and the same measure target population. This measure is the most valid and efficient for capturing UTI among nursing home residents for purposes of improving genitourinary healthcare quality and resident safety in this domain.

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) Percent of Residents with a Urinary Tract Infection (Long Stay) (NQF #0684); and (2) Percent 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: 05, 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.