**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](http://www.qualityforum.org/Measuring_Performance/Submitting_Standards.aspx). * 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**](#Note10) 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**](#Note11) 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**](#Note12)  **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**](#Note13)  **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**](#Note14)**,**[**15**](#Note15) 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**](#Note16) **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.
   1. 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]](#footnote-2)

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. 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* ≥ 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% |
| 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% |
| 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., *ρ* = ). 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 ( = 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 (=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* ≤ .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.

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***](#section2b4)

**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***](#section2b5)***.***

**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 (*χ*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 (*χ*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 (*χ*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 (*χ*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 chi2 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% |
| 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 variationin the social risk factor among nursing homes (there is; see ***Table 6***), and the potential improvement in the risk modelif 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 Chi2, *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***](#question2b49)

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 chi2 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.

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. [↑](#footnote-ref-2)