

October 2016

Hospice Quality Reporting Program: HIS-Based Quality Measures Annual Testing Executive Summary

Prepared for

Center for Clinical Standards and Quality
Centers for Medicare & Medicaid Services
7500 Security Boulevard
Baltimore, MD 21244-1850

CMS Contract No. HHSM-500-2013-13015I



HOSPICE QUALITY REPORTING PROGRAM:
HIS-BASED QUALITY MEASURES ANNUAL TESTING EXECUTIVE SUMMARY

RTI International

CMS Contract No. HHSM-500-2013-13015I

October 2016

This project was funded by the Centers for Medicare & Medicaid Services under contract no. HHSM-500-2013-13015I. The statements contained in this report are solely those of the authors and do not necessarily reflect the views or policies of the Centers for Medicare & Medicaid Services. RTI assumes responsibility for the accuracy and completeness of the information contained in this report.

EXECUTIVE SUMMARY

Section 3004 of the Patient Protection and Affordable Care Act (ACA) authorizes the Health and Human Services Secretary to establish a quality reporting program for hospices. The ACA specifies that for fiscal year (FY) 2014 and each subsequent FY, hospice programs shall submit to the Secretary data on quality measures (QMs). The ACA also describes measure endorsement requirements for any measures specified by the Secretary. The Centers for Medicare & Medicaid Services (CMS) implemented the Hospice Quality Reporting Program (HQRP) in the FY 2012 Hospice Wage Index final rule (76 FR 47302-47352). CMS implemented the Hospice Item Set (HIS), a standardized, patient-level data collection instrument, as part of the HQRP in the FY 2014 Hospice Wage Index final rule (78 FR 48234-48281). Medicare-certified hospices are required to submit an HIS-Admission record and an HIS-Discharge record for each patient admission on or after July 1, 2014.

The current version of the HIS (V1.00.1) collects individual-level data to calculate seven National Quality Forum (NQF)-endorsed QMs. These measures focus on care processes around hospice admission that are either clinically recommended, required in the hospice Conditions of Participation, or both. The current seven HQRP QMs are described in *Table 1*.

Table 1
Description of the Seven HQRP QMs

Measure Title (NQF ID)	Measure Description
Treatment Preferences (NQF #1641)	The percentage of hospice patient stays with chart documentation that the hospice discussed (or attempted to discuss) preferences for life-sustaining treatments.
Beliefs/Values Addressed (if desired by the patient) (NQF #1647)	The percentage of hospice patient stays with documentation of a discussion of spiritual and existential concerns or documentation that the patient and/or caregiver did not want to discuss.
Pain Screening (NQF #1634)	The percentage of hospice patient stays during which the patient was screened for pain during the initial nursing assessment.
Pain Assessment (NQF #1637)	The percentage of hospice patient stays during which the patient screened positive for pain and received a comprehensive assessment of pain within 1 day of screening.
Dyspnea Screening (NQF #1639)	The percentage of hospice patient stays during which the patient was screened for dyspnea during the initial nursing assessment.
Dyspnea Treatment (NQF #1638)	The percentage of hospice patient stays during which the patient screened positive for dyspnea and received treatment within 1 day of the screening.
Patients Treated with an Opioid who are Given a Bowel Regimen (NQF #1617)	The percentage of patient stays with vulnerable adults treated with an opioid that are offered/prescribed a bowel regimen or documentation of why this was not needed.

To test the scientific soundness of the QMs, we conducted analyses that assessed the overall performance of the QMs using the criteria specified in the guidelines for determining measure reliability and validity set forth in the CMS Measures Management System Blueprint

and by the NQF. These analyses addressed the following key areas: reportability, distribution and variability, reliability, and validity. We also examined disparities in hospice care.

We conducted analysis on hospice patient stays discharged from October 1, 2014, through March 31, 2016. We used both HIS records from July 1, 2014, through March 31, 2016, as well as the Provider of Services File (containing hospice characteristics updated as of their latest survey at the end of June 2015). The HQRP QMs are calculated and reported using four consecutive quarters of data. Therefore, we were able to perform analysis on three reporting periods: Q4 2014–Q3 2015, Q1–Q4 2015, and Q2 2015–Q1 2016. Below, we describe the analytic approaches and summarize the findings. The results are largely consistent across the three reporting periods; therefore, the presentation and summary of results focuses on the findings from the latest reporting period, Q2 2015–Q1 2016.

Hospice and Patient Characteristics

We conducted descriptive analyses to examine the hospice characteristics, including hospices' original Medicare participation/certification year, ownership type, facility type, and urban/rural status. Although 3,971 hospices submitted HIS data from Q2 2015 through Q1 2016, we did not have characteristics for 81 of these facilities. Approximately 14 percent of hospices were certified in the 1980s, 27 percent in the 1990s, 33 percent in the 2000s, and 24 percent in 2010 or later. Most hospices were for-profit (almost 59%), and about 25 percent of hospices were nonprofit. About 14 percent of hospices were owned by the government or other entities, and 76 percent of hospices were freestanding, meaning they were not part of a hospital, skilled nursing facility, or home health agency. Most hospices were located in urban areas (about 75%).

Using the HIS data, we conducted descriptive analyses to examine patient characteristics, including demographics, diagnosis, site of service at admission, and where the patient was admitted from. From Q2 2015 through Q1 2016, there were 1,298,610 patient stays, and the average age of patients on the day of admission was 78. About 34 percent of patients had a principal diagnosis of cancer, and almost 14 percent had a principal diagnosis of dementia/Alzheimer's. Over half of patients (about 52%) received hospice care in their homes or residences. Most patients were admitted from either a community or residential setting (about 44%) or from a short-stay acute hospital (about 33%). The most common reason for hospice discharge was patient expiration (about 84%), but a nontrivial proportion (about 16%) of patient stays were discharged alive for various reasons, including if the patient revoked the hospice benefit, was no longer terminally ill, moved out of the hospice service area, was transferred to another hospice, or was discharged for cause.

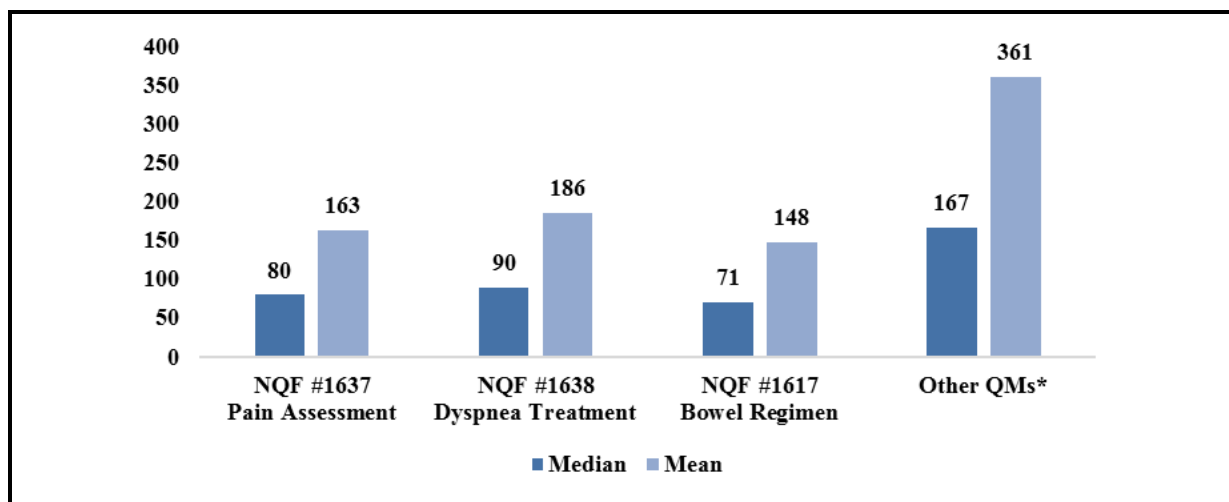
Reportability

Reportability analysis assesses whether the sample size for each QM is large enough to generate statistically reliable scores. Without sufficient sample size, a hospice's QM score can be significantly affected by outliers or extreme values, yielding scores that are not statistically reliable. Publicly displaying QM scores on the basis of small samples can potentially give users of public reporting misleading information about quality of care. Therefore, CMS set a minimum sample size of 20 qualifying stays to ensure that hospices with small sample sizes will have their scores suppressed from public display. For this analysis and subsequent analyses presented in this report, we only included hospices that met the minimum sample size threshold. We

examined the impact of this threshold on the number of hospices that would qualify for the public reporting of each QM. We found that applying the minimum threshold would suppress about 10 percent of hospices from public reporting on four out of the seven QMs, and about 19–28 percent from public reporting on the other three QMs. **Figure 1** presents the mean and median sample sizes of the QMs in Q2 2015–Q1 2016.

Four of the QMs (*NQF #1641, Treatment Preferences; NQF #1647, Beliefs/Values; NQF #1634, Pain Screening; and NQF #1639, Dyspnea Screening*) had a mean sample size of about 361 stays and a median sample size of 167 stays. Mean sample sizes for the other three QMs ranged from about 148 to 186 stays, and the median sample sizes for these QMs ranged from 71 to 90 stays. For all QMs, the mean sample size is larger than the median because there were hospices with large sample sizes.

Figure 1
Median and Mean Sample Sizes for HQRP QMs



SOURCE: RTI International analysis of HIS data (April 2015–March 2016).

NOTE: Results in the figure are rounded to the nearest patient stay.

*Four QMs–NQF #1641, Treatment Preferences; NQF #1647, Beliefs/Values; NQF #1634, Pain Screening; and NQF #1639, Dyspnea Screening–have the same sample size.

Distribution of QM Scores and Variability

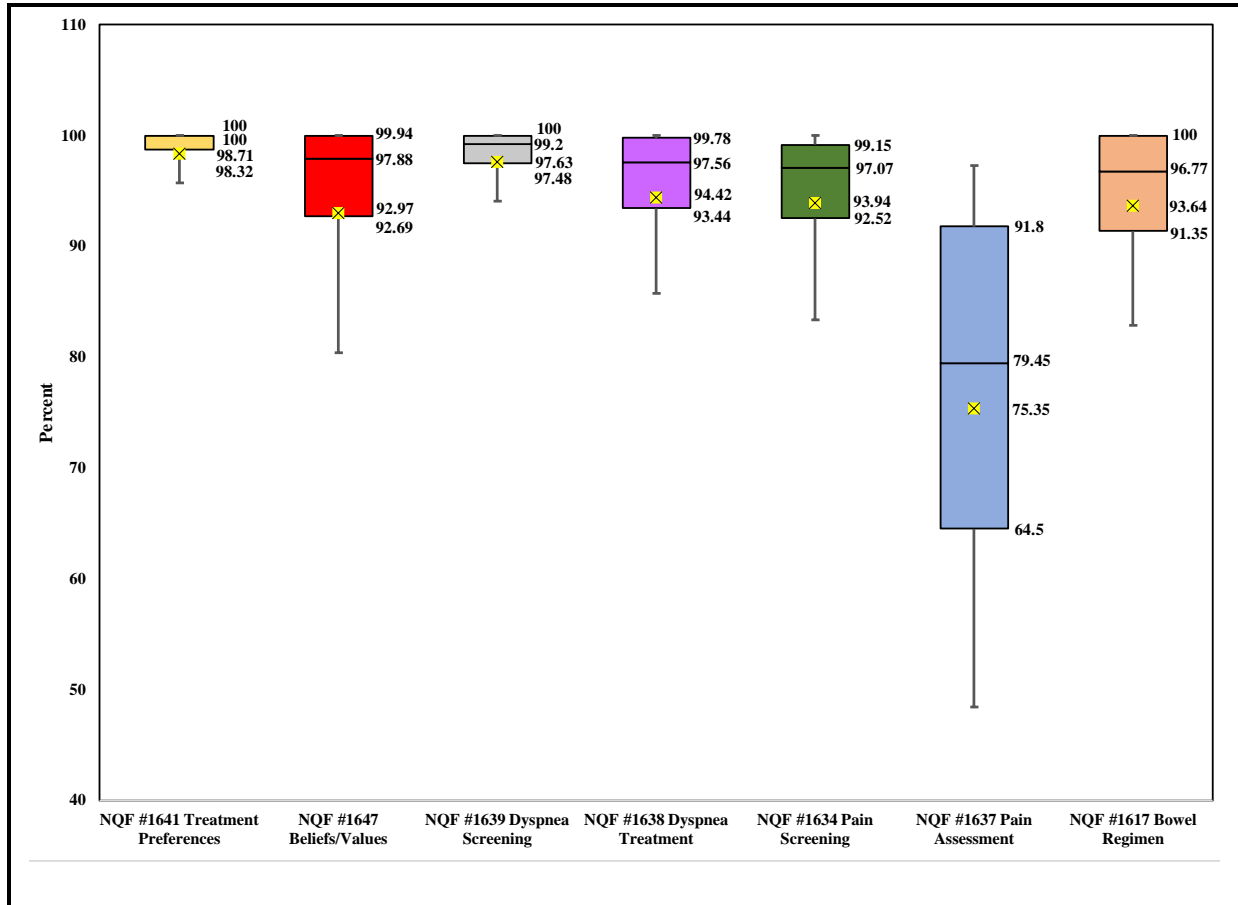
Variability refers to the spread or the differences in QM scores across facilities. A meaningful and useful QM should have sufficient variability across providers to distinguish between high- and low-quality hospices. We assessed variability by analyzing the distribution of hospice-level scores. **Figure 2** presents the distributions of the seven QMs in Q2 2015–Q1 2016.

- The QM scores were generally high for six of the seven QMs: *NQF #1641, Treatment Preferences; NQF #1647, Beliefs/Values; NQF #1634, Pain Screening; NQF #1639, Dyspnea Screening; NQF #1638, Dyspnea Treatment; and NQF #1617, Bowel Regimen*. The national mean scores for these QMs (represented by Xs in **Figure 2**) ranged from 92.97 percent for *NQF #1647, Beliefs/Values*, to 98.32 percent for *NQF #1641, Treatment Preferences*. In other words, hospices completed these care

processes upon admission for about 93 to 98 percent of their patients, on average. The median ranged from 96.77 percent for *NQF #1617, Bowel Regimen*, to 100 percent for *NQF #1641, Treatment Preferences*. Compared with these six QMs, hospices' performance on *NQF #1637, Pain Assessment*, was lower, with a national mean of 75.35 percent and median of 79.45 percent. On average, hospices completed a pain assessment for about 75 percent of their patients that screened positive for pain upon admission.

- Six of the seven QMs had relatively small variability, as indicated by the interquartile range (the boxes in *Figure 2*). The QM with the smallest variability was *NQF #1641, Treatment Preferences*, with an interquartile range of only 1.29 percent, suggesting that a large proportion of hospices had similar QM performance. *NQF #1637, Pain Assessment*, had the largest variability, with an interquartile range of 27.30 percent.

Figure 2
Distribution of Hospice-Level Scores on HQRP QMs



SOURCE: RTI International analysis of HIS data (April 2015–March 2016).

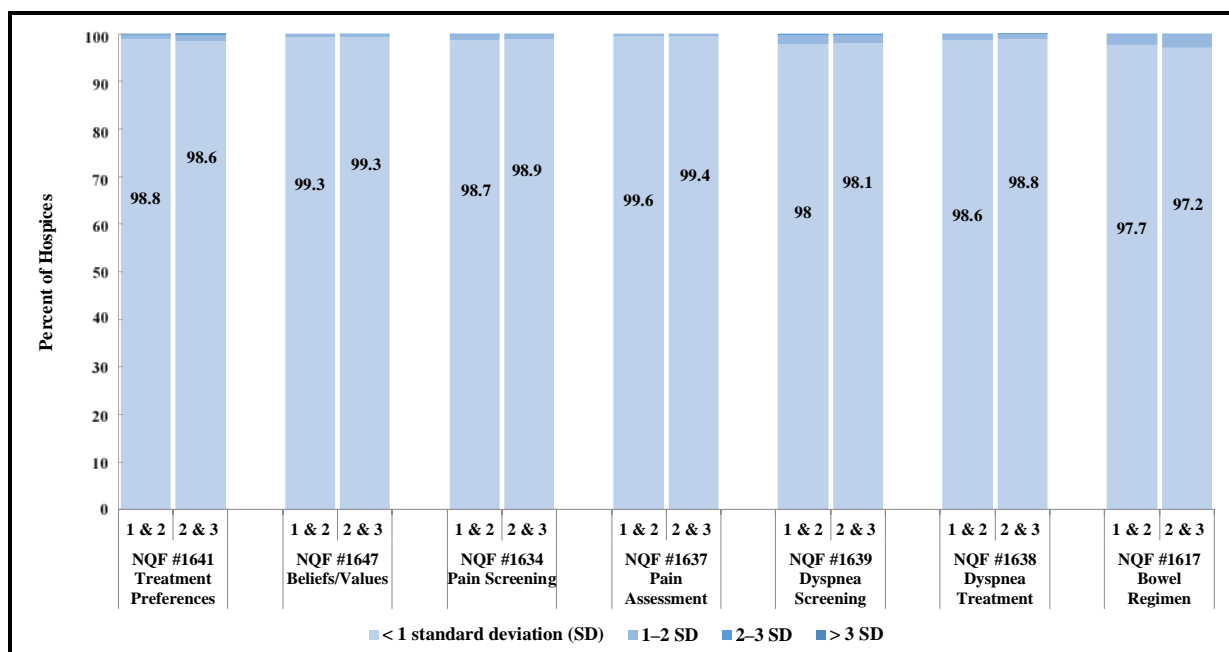
NOTE: The boxes in the figure represent the interquartile range, and horizontal bars within the box represent the median scores. The Xs are the mean scores. The end bars out of the boxes represent the 10th and 90th percentile QM scores. For *NQF #1641, Treatment Preferences*, the mean was 98.32%, slightly lower than the 25th percentile, 98.71%.

Reliability

Reliability analyses assess the extent to which a QM produces consistent results about the quality of care. Reliability is a key scientific metric of a QM because it describes how well a measure can reliably distinguish the performance of one provider from another. We conducted three types of analyses to test the reliability of the QMs: stability analysis, split-half analysis, and signal-to-noise ratio analysis. All three analyses indicated high reliability of these QMs.

- Stability analysis.** Stability analysis describes the extent to which providers' performance assessed by a QM changes across time. Below, **Figure 3** illustrates the change in facility scores between reporting periods. The changes in facility scores are reported in standard deviations. Over 97 percent of facilities had a change in QM score of less than one standard deviation, indicating high stability of the QMs.

Figure 3
Stability: Standardized QM Score Changes Across Reporting Periods



SOURCE: RTI International analysis of HIS data (October 2014–March 2016).

NOTE: For each QM, the left column presents the standardized score change between the first and second reporting periods (1 & 2: Q4 2014–Q3 2015 vs. Q1 2015–Q4 2015), and the right column presents the change between the second and third reporting periods (2 & 3: Q1 2015–Q4 2015 vs. Q2 2015–Q1 2016).

- Split-half analysis.** Split-half analysis assesses the internal consistency of a QM by randomly dividing the patient stays within each hospice into two halves and calculating the correlation between the hospice's QM scores on the basis of the two randomly divided halves. In this analysis, we used the intraclass correlation (ICC) coefficients to measure the internal reliability. For the most recent reporting period, Q2 2015–Q1 2016, the ICC coefficients of the QMs ranged from 0.77 for *NQF #1617, Bowel Regimen*, to 0.94 for *NQF #1647, Belief/Values*, indicating high

internal reliability (*Table 2*). The ICC coefficients for each QM remained relatively stable across the reporting periods.

Table 2
Reliability: Split-Half Reliability and Signal-to-Noise Analysis

QMs	ICC Coefficient	Signal-to-Noise Ratio
NQF #1641, Treatment Preferences	0.90	0.97
NQF #1647, Beliefs/Values	0.94	0.99
NQF #1634, Pain Screening	0.87	0.98
NQF #1637, Pain Assessment	0.89	0.98
NQF #1639, Dyspnea Screening	0.83	0.98
NQF #1638, Dyspnea Treatment	0.82	0.96
NQF #1617, Bowel Regimen	0.77	0.95

SOURCE: RTI International analysis of HIS data (April 2015–March 2016).

- *Signal-to-noise analysis.* If a measure is reliable, then true differences in provider performance, rather than randomly distributed variation, should explain a substantial proportion of the variance in QM scores. We conducted an analysis of variance to determine what proportion of total variance in the QM scores was attributable to differences among providers. For the most recent reporting period, Q2 2015–Q1 2016, the results showed that more than 97 percent of the variance in the QMs was attributable to differences among facilities, indicating high reliability (*Table 2*). The signal-to-noise ratio analysis for each QM across multiple reporting quarters consistently revealed high reliability of the measures.

Validity

Validity refers to whether a QM captures the actual quality of care that it is intended to measure. We conducted correlation analysis of the seven QMs to assess validity. The results provide some evidence to support validity of the measures.

- *Correlations.* Providers should perform similarly on QMs that reflect similar care processes and similar patient populations. We conducted nonparametric Spearman rank correlation analysis among all the hospices' ranks on seven HQRP QMs. Our results showed that the Spearman correlation coefficients were positive and statistically significant. The significant positive correlations between every pair of QMs indicated that hospices that provided high-quality care around hospice admission in one area also provided high-quality care in other areas. For example, hospices that perform better with respect to discussion of treatment preferences also provided higher quality of care regarding discussion of patients' beliefs/values. Overall, the correlations between the QMs were low to moderate, ranging from 0.06 to 0.64. The skewed distributions and low variability across all seven HQRP QMs can affect the level of correlations between QMs. For example, the low correlation (0.06) between *NQF #1637, Pain Assessment*, and *NQF #1638, Dyspnea Screening*, could

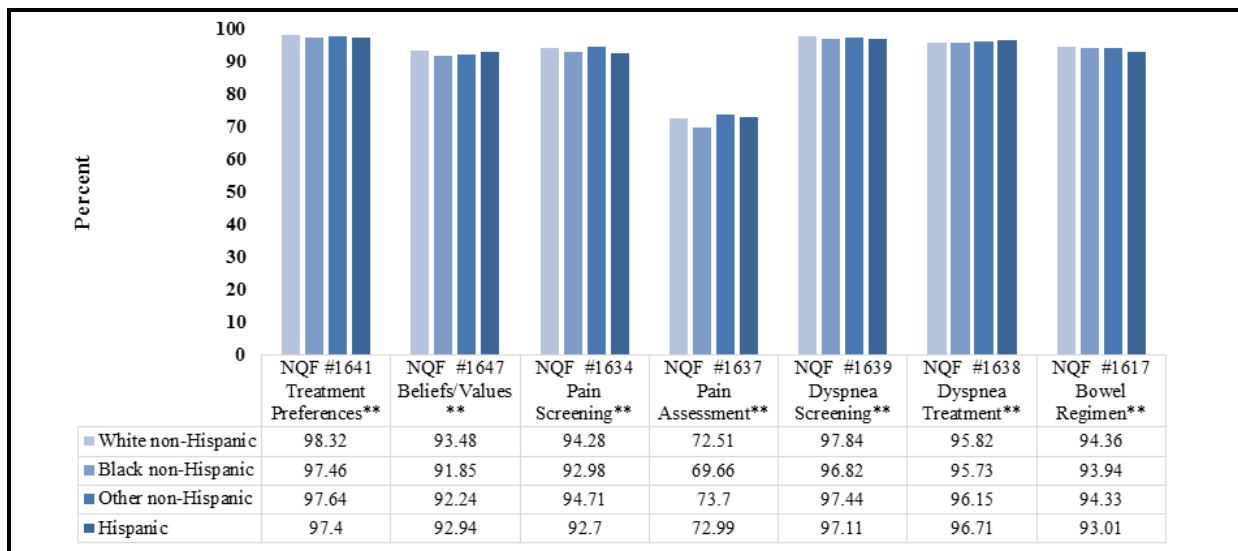
be a result of the vast majority of hospices performing highly on the dyspnea screening measure while scores for the pain assessment measure were generally lower and much more variable.

Disparity Analysis

We analyzed disparities in QM scores based on patient and hospice characteristics. Patient characteristics included race or ethnicity and gender. Hospice characteristics included the proportion of the patient population that was nonwhite, the proportion of the patient population that was female, and the location of the hospice (urban or rural). At the patient level, we compared the QM scores across sociodemographic groups. At the hospice level, we compared the QM scores between facilities with varying population sociodemographic characteristics. We found small but statistically significant differences in QM scores between patients and hospices with different characteristics.

- Racial and ethnic disparity.* **Figure 4** presents the QM scores for the four racial/ethnic groups: white non-Hispanic, black non-Hispanic, other non-Hispanic, and Hispanic. Overall, we found small differences in the QM scores across the four groups. Compared with white non-Hispanic and other non-Hispanic patients, a slightly smaller proportion of the Hispanic and black non-Hispanic patients received the care processes measured by four of the seven QMs: *NQF #1641, Treatment Preferences; NQF #1634, Pain Screening; NQF #1639, Dyspnea Screening; and NQF #1617, Bowel Regimen.* Over the course of three consecutive reporting periods, the variation in each QM by race and ethnic group remains small but statistically significant.

Figure 4
Differences in QM Scores between Four Racial/Ethnic Groups

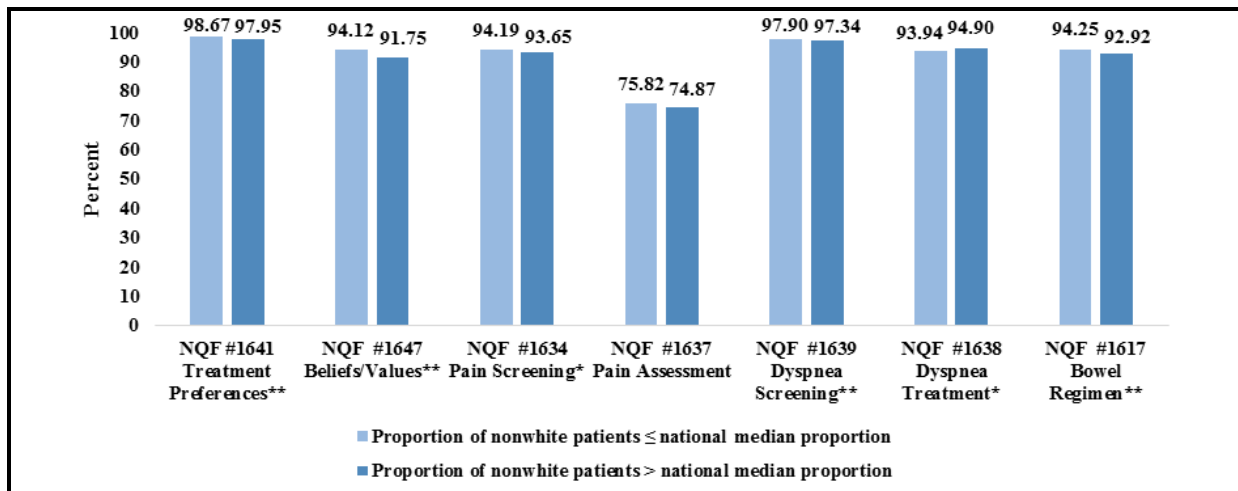


SOURCE: RTI International analysis of HIS data (April 2015–March 2016).

** $P < 0.01$.

The hospice-level analysis, presented in **Figure 5**, indicated that hospices that serve fewer nonwhite patients (i.e., with a proportion of nonwhite patients smaller than the national median) have statistically significantly higher scores on five QMs: *NQF #1641, Treatment Preferences* (98.67% vs. 97.95%); *NQF #1647, Beliefs/Values* (94.12% vs. 91.75%); *NQF #1634, Pain Screening* (94.19% vs. 93.65%); *NQF #1639, Dyspnea Screening* (97.90% vs. 97.34%); and *NQF #1617, Bowel Regimen* (94.25% vs. 92.92%). For one measure, *NQF #1638, Dyspnea Treatment*, this relationship was reversed and hospices with a higher proportion of nonwhite patients had statistically significantly higher scores (94.90% vs. 93.94%).

Figure 5
Mean Hospice QM Scores by Proportion of Nonwhite Patients



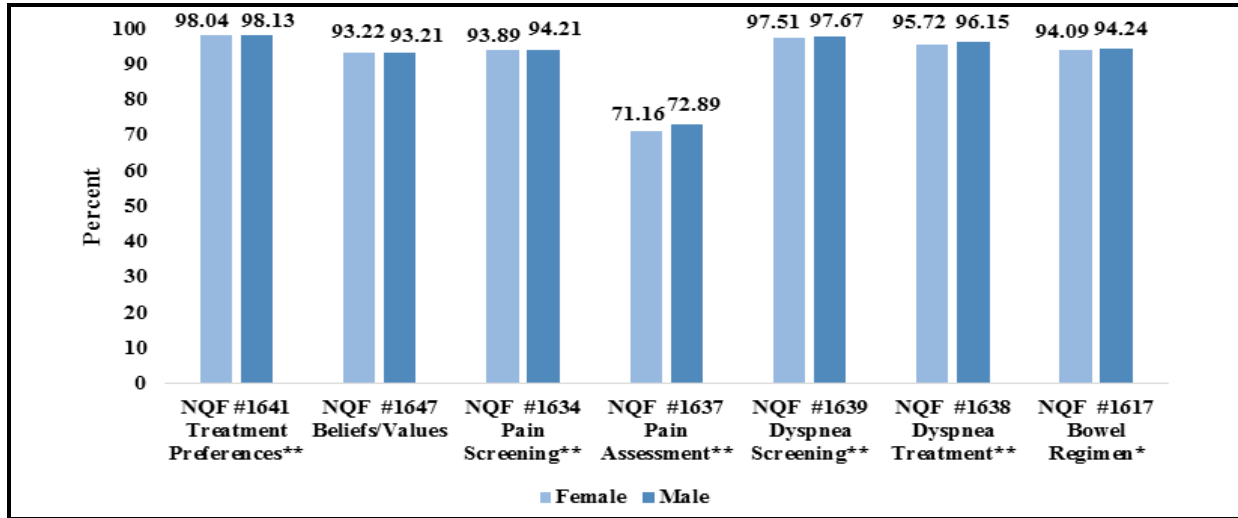
SOURCE: RTI International analysis of HIS data (April 2015–March 2016).

NOTE: The national median proportion of nonwhite patients is 12.96 percent. The median calculation includes all hospices regardless of denominator size.

* $P < 0.05$. ** $P < 0.01$.

- Gender disparity.** The patient stay-level analysis, presented in **Figure 6**, indicated small gender disparity in some care processes. A smaller proportion of female patients received these six care processes: discussion of treatment preferences (98.04% vs. 98.13%), pain screening (93.89% vs. 94.21%), pain assessment (71.16% vs. 72.89%), dyspnea screening (97.51% vs. 97.67%), dyspnea treatment (95.72% vs. 96.15%), and bowel regimen (94.09% vs. 94.24%). Although statistically significant, these differences were small. For example, for the pain assessment measure, approximately 71 out of 100 women who screened positive for pain received a full pain assessment, compared with 73 out of 100 men. There is no statistically significant difference in whether female and male patients were asked about beliefs and values.

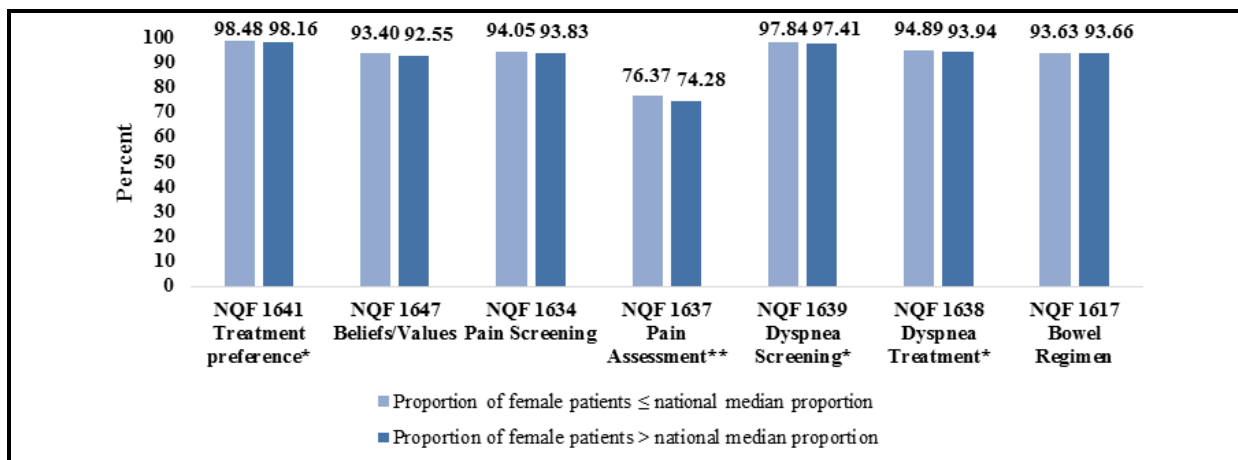
Figure 6
Differences in QM Scores Between Female and Male Patients



SOURCE: RTI International analysis of HIS data (April 2015–March 2016).
 * $P < 0.05$. ** $P < 0.01$.

The hospice-level analysis, presented in **Figure 7**, indicated that hospices that provide care to more female patients (i.e., with a proportion of female patients greater than the national median) have mean hospice-level scores that are statistically significantly lower on four QMs: *NQF #1641, Treatment Preferences* (98.16% vs. 98.48%); *NQF #1637, Pain Assessment* (74.28% vs. 76.37%); *NQF #1639, Dyspnea Screening* (97.41% vs. 97.84%); and *NQF #1638, Dyspnea Treatment* (93.94% vs. 94.89%). No statistically significant difference was found for the other three QMs.

Figure 7
Mean Hospice QM Scores by Proportion of Female Patients



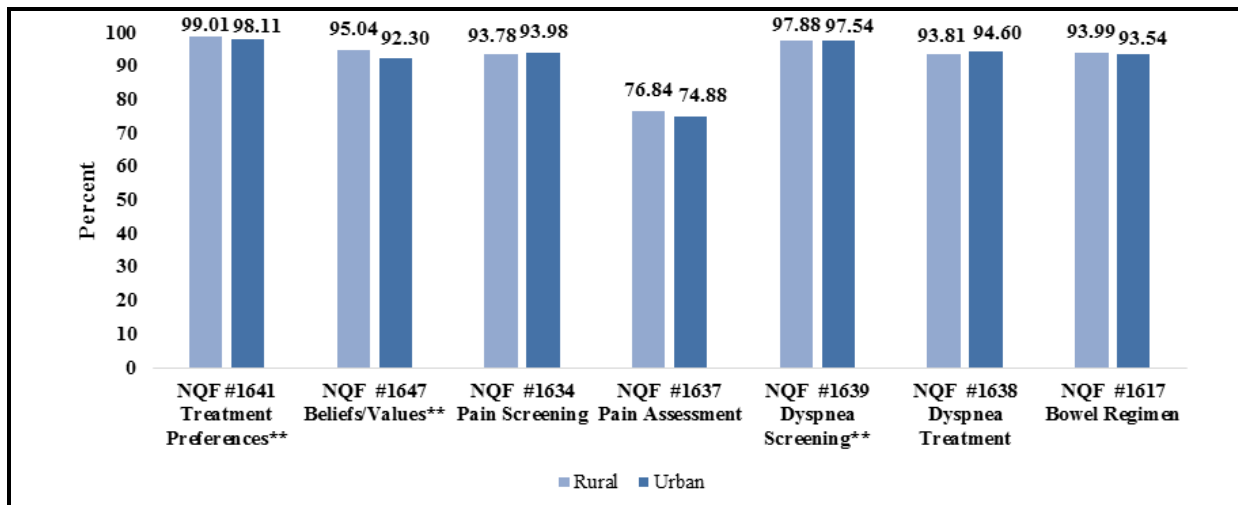
SOURCE: RTI International analysis of HIS data (April 2015–March 2016).

NOTE: The national median proportion of female patients is 55.72 percent. The median calculation includes all hospices regardless of denominator size.

* $P < 0.05$. ** $P < 0.01$.

- Rural vs. urban. **Figure 8** presents the mean QM scores of rural and urban hospices. Compared with urban hospices, rural hospices had statistically significantly higher scores on three QMs, indicating better care: *NQF #1641, Treatment Preferences* (99.01% vs. 98.11%); *NQF #1647, Beliefs/Values* (95.04% vs. 92.30%); and *NQF #1639, Dyspnea Screening* (97.88% vs. 97.54%). The differences between rural and urban hospices were not significant on *NQF #1634, Pain Screening* (93.78% vs. 93.98%), *NQF #1637, Pain Assessment* (76.84% vs. 74.88%), *NQF #1638, Dyspnea Treatment* (93.81% vs. 94.60%), and *NQF #1617, Bowel Regimen* (93.99% vs. 93.54%).

Figure 8
Mean Hospice-Level QM Scores in Rural and Urban Areas



SOURCE: RTI International analysis of HIS data (April 2015–March 2016).

* $P < 0.05$. ** $P < 0.01$.

Summary of Findings

Our analyses demonstrate that the current seven HQRP QMs perform well on various tests of measure performance. Thus, they provide reliable, valid, and useful information about quality of hospice care.

- Reportability.** With the minimum sample size threshold applied, about 90 percent of hospices would qualify for public reporting on four out of the seven QMs, and about 72–80 percent would be included in the public reporting of the other three QMs. Mean denominator sizes ranged from 148 to 361 stays, and median denominator sizes ranged from 71 to 167 stays.
- Distribution and variability.** Hospices are providing the required and recommended care to the majority of the patients around hospice admission, demonstrating overall high quality of care. There is larger variation across hospices in providing standardized pain assessment to patients who screened positive for pain symptoms. Among the currently assessed and reported care processes in HQRP, pain assessment is the one that can be best used to differentiate hospices on the basis of the quality of

care. At the same time, hospices have a larger opportunity to improve their performance in this area.

- *Reliability.* The current seven HQRP QMs have high reliability, demonstrated by high stability, internal consistency, and signal-to-noise ratio.
- *Validity.* Hospices that are high-performing on one care process also provide high-quality care on other care processes around hospice admission. Overall, the correlations between the QMs are low to moderate. Although the correlations of QMs for other care settings are also generally low, the skewed distributions and low variability across all seven HQRP QMs can affect the level of correlations between QMs. Thus, the low correlations identified do not necessarily indicate low validity, and the overall statistically significant correlations support the overall validity of the measure set.
- *Disparities.* We found statistically significant but small differences in QM scores across patients and hospices with different sociodemographic characteristics.