**Summary of changes:**

Currently the ”Heart Failure Symptoms Addressed” measure assesses whether the clinician addressed a patient’s symptoms of heart failure when the clinician assessed the patient to have symptoms of heart failure. We have proposed revisions to the measure specifications so that agencies will be held accountable for assessing heart failure symptoms in all patients with a diagnosis of heart failure, as well as addressing those symptoms when they are present. In addition, the measure now applies to both short-term and long-term home health care episodes - long-term home health care episodes are no longer excluded.

Table : Comparison of Revised and Current Specifications for Measure 0521

|  |  |  |
| --- | --- | --- |
| Measure Components | Proposed Specifications | Current Specifications |
| Description | Percentage of home health episodes of care during which patients with heart failure were assessed for symptoms of heart failure, and appropriate actions were taken when the patient exhibited symptoms of heart failure . | Percentage of home health episodes of care during which patients exhibited symptoms of heart failure and appropriate actions were taken (since the previous OASIS assessment). |
| Numerator | Number of home health episodes of care during which patients with heart failure were assessed for symptoms of heart failure and appropriate actions were taken when the patient exhibited symptoms of heart failure. | Number of home health episodes of care during which patients exhibited symptoms of heart failure and appropriate actions were taken (since the previous OASIS assessment). |
| Denominator | Number of home health episodes of care ending with a discharge or transfer to inpatient facility during the reporting period for patients with a diagnosis of heart failure, other than those covered by generic or measure-specific exclusions. | Number of home health episodes of care ending with a discharge or transfer to inpatient facility during the reporting period, other than those covered by generic or measure-specific exclusions. |
| Measure-specific Exclusions | Home health episodes for which patient does not have heart failure diagnosis OR patient died. | Home health episodes for which patient does not have heart failure diagnosis, OR heart failure symptoms were not assessed, OR no heart failure symptoms exhibited since the previous assessment, OR patient died. |
| OASIS C Item(s) Used | (M0100) Reason for Assessment  (M1020) Primary Diagnosis and ICD Code\*  (M1022) Other Diagnoses and ICD Codes\*  (M1500) Symptoms in Heart Failure Patients  (M1510) Heart Failure Follow-up | (M0100) Reason for Assessment  (M1500) Symptoms in Heart Failure Patients  (M1510) Heart Failure Follow-up |

\* See Table 8 for a list of the relevant ICD-9 and ICD-10 codes

Using existing specifications, patients with heart failure are identified as those having a response (other than “N/A”) on OASIS item M1500. Although M1500 (Symptoms in Heart Failure Patients) is applicable for all heart failure patients, recent analyses conducted by Acumen suggest that a substantial number of patients marked “N/A – Patient does not have a diagnosis of heart failure” on item M1500 actually have a diagnosis of heart failure indicated on M1020 or M1022. The revised specifications

With the proposed changes, the measure denominator will be defined as all home health episodes for patients with heart failure. Patients with heart failure are identified based on having a response (other than “N/A”) on M1500 and/or having ICD-9 codes corresponding to heart failure (see Attachment B ) included on M1020 (i.e., primary home health diagnosis) or M1022 (i.e., additional home health diagnoses). Combining information from all three items is necessary to correctly identify patients with heart failure.

For all home health episodes in the measure denominator, the proposed measure logic then considers whether symptoms were found to be present at any point since the previous OASIS assessment (M1500). If M1500 indicates no symptoms, the value of Heart Failure Symptoms Addressed is set to 1. If M1500 indicates symptoms, the logic then considers whether any action was taken during follow-up (M1510). M1510 values affirming follow-up action will set the measure value to 1, while no indication of action taken will set the measure to 0. Finally, if M1500 takes a value of 2 (Not Assessed) or NA (Patient does not have a diagnosis of heart failure), the measure is set to 0.

**Opportunity for Improvement**

The measure, as currently specified, holds agencies accountable only for addressing symptoms in patients who exhibit symptoms of heart failure. Using OASIS assessments from July 1, 2012 through June 30, 2013, we calculated home health agencies’ performance rates for the measure using the existing specifications and found high levels of performance with little variation in the measure scores across agencies and therefore, little room for improvement. As shown in Table 3 of this attachment, the average agency rate with the current specifications is **98.0** percent across agencies with at least 20 valid episodes of care.

We are proposing changes to the measure so that agencies will be held accountable both for assessing heart failure symptoms in all patients with a diagnosis of heart failure, in addition to addressing those symptoms. These proposed changes raise the quality standards required to meet the measure criteria and this is reflected in a greater opportunity for improvement for this measure.

Using OASIS assessments from July 1, 2012 through June 30, 2013, Acumen calculated the measure rates using the revised specifications, and found evidence that the revision to the measure specifications would increase the opportunity for agencies to improve their care delivery for patients at risk of heart failure.

Table 2 shows the distribution of observed rates and numerator and denominator counts of the *Heart Failure Symptoms Addressed* measure under the revised and current measure specifications, by population group. Overall, implementing the revised specification causes the average performance rate to decrease from 98.0 percent to 92.6 percent among all patients, and by several percentage points across all population groups. The denominator under the revised specifications is larger across all population groups because it includes all patients with a diagnosis of heart failure and with symptoms assessed, compared to the current inclusion criteria of all patients who exhibited symptoms since the previous assessment. There were no apparent differences in the effect of revising the specification logic between groups within the population categories of race, age, and region groups. However, for the agency size category, we observed a relatively larger decrease in measure performance for patients in large agencies compared to patients in smaller agencies as a result of the proposed. Further investigation into the factors associated with this difference may be warranted.

Table : Observed Rates of *Heart Failure Symptoms Addressed* by Population Group

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Population Group | | Observed Rate | | Numerator | | Denominator | |
| **Revised**  **Logic** | **Current Logic** | **Revised Logic** | **Current Logic** | **Revised**  **Logic** | **Current Logic** |
| **All Patients** | | 92.6% | 98.0% | 1,789,575 | 336,867 | 1,932,296 | 343,572 |
| **By Race** | White | 92.4% | 98.1% | 1,319,889 | 263,729 | 1,427,739 | 268,962 |
| Black | 92.6% | 98.0% | 271,451 | 50,469 | 293,187 | 51,513 |
| Hispanic | 93.7% | 98.1% | 154,443 | 16,114 | 164,816 | 16,420 |
| Other | 94.1% | 98.2% | 43,792 | 6,555 | 46,554 | 6,677 |
| **By Age** | <65 | 93.4% | 97.8% | 288,478 | 44,789 | 308,991 | 45,777 |
| 65-75 | 93.0% | 98.0% | 429,233 | 75,490 | 461,323 | 77,020 |
| 75-85 | 92.6% | 98.1% | 572,046 | 111,362 | 617,938 | 113,510 |
| 85+ | 91.9% | 98.1% | 499,818 | 105,226 | 544,044 | 107,265 |
| **By Gender** | Male | 92.6% | 98.1% | 701,223 | 140,870 | 757,118 | 143,591 |
| Female | 92.6% | 98.0% | 1,088,352 | 195,997 | 1,175,178 | 199,981 |
| **By Agency Size\*** | Very Small | 96.2% | 98.2% | 8,947 | 1,334 | 9,296 | 1,358 |
| Small-Medium | 94.9% | 97.7% | 134,838 | 19,506 | 142,051 | 19,956 |
| Large | 92.4% | 98.1% | 1,645,790 | 316,027 | 1,780,949 | 322,258 |
| **By Region\*\*** | Region 1 | 93.3% | 98.2% | 112,825 | 24,668 | 120,866 | 25,111 |
| Region 2 | 92.6% | 98.2% | 171,070 | 25,553 | 184,672 | 26,030 |
| Region 3 | 91.9% | 98.2% | 184,308 | 37,300 | 200,444 | 37,997 |
| Region 4 | 92.4% | 97.7% | 438,524 | 68,074 | 474,534 | 69,647 |
| Region 5 | 93.0% | 98.3% | 328,926 | 78,496 | 353,542 | 79,820 |
| Region 6 | 92.3% | 98.0% | 216,583 | 42,799 | 234,739 | 43,675 |
| Region 7 | 92.7% | 97.9% | 67,387 | 15,580 | 72,661 | 15,922 |
| Region 8 | 91.6% | 96.8% | 35,509 | 6,513 | 38,785 | 6,728 |
| Region 9 | 93.2% | 98.4% | 189,295 | 29,206 | 203,167 | 29,685 |
| Region 10 | 92.4% | 96.9% | 45,148 | 8,678 | 48,886 | 8,957 |

*\*Very Small = <20 Episodes, Small-Medium = >20 and less than the median # of episodes, Large = >=Median # of episodes \*\*Region refers to CMS region*

Table 3 shows the distribution of observed agency rates for the *Heart Failure Symptoms Addressed* measure. Across agencies with at least 20 valid episodes of care, the revision to the measure specifications would decrease the average performance rate on the measure by approximately 5 percentage points, (i.e., from 98.0 percent to 92.9 percent). Additionally, for agencies with at least 20 valid episodes of care, the interquartile range of the measure rates under the proposed specifications is 6.9 (i.e., 97.4 - 90.5); an agency at the 75th percentile still has a measure rate that is 6.9 percentage points higher than an agency at the 25th percentile. In contrast, under the current specifications, the corresponding interquartile range is only 3.0 (100.0 – 97.0). Performance among agencies in the lowest decile falls from 95.0 percent to 85.0 percent, again showing that the revised specification allows more room for improvement. Finally, the number of agencies with at least 20 valid episodes of care increases from only 3,618 under the current measure logic to 8,882 under the revised logic, expanding public reporting eligibility.

Table : Distribution of Observed Agency Rates of *Heart Failure Symptoms Addressed*

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Measure Specifications | | # of Agencies | Agency Avg. Rate | Percentile | | | | | | |
| **min** | **p10** | **p25** | **p50** | **p75** | **p90** | **max** |
| **All Agencies** | Current Logic | 11,849 | 95.9% | 0.0% | 93.0% | 99.0% | 100.0% | 100.0% | 100.0% | 100.0% |
| Revised Logic | 11,849 | 92.5% | 0.0% | 83.3% | 90.5% | 95.2% | 100.0% | 100.0% | 100.0% |
| **Agencies with > 20 Valid Episodes** | Current Logic | 3,618 | 98.0% | 59.0% | 95.0% | 97.0% | 100.0% | 100.0% | 100.0% | 100.0% |
| Revised Logic | 8,882 | 92.9% | 0.7% | 85.0% | 90.5% | 94.6% | 97.4% | 100.0% | 100.0% |

# Measure Reliability

Acumen conducted two types of reliability testing on the revised measure, including calculating, at the agency level, the (i) beta binomial reliability scores, and (ii) the intraclass correlation coefficient (ICC). The remainder of this section describes each in turn.

## Beta Binomial Reliability Scores

To address the reliability of the performance scores, Acumen measured the extent to which differences in each quality measure were due to actual differences in agency performance versus variation that arises from measurement error. Statistically, reliability depends on performance variation for a measure across agencies, the random variation in performance for a measure within an agency’s panel of attributed beneficiaries, and the number of beneficiaries attributed to the agency. High reliability for a measure suggests that comparisons of relative performance across agencies are likely to be stable over different performance periods, and that the performance of one agency on the quality measure can confidently be distinguished from another. Potential reliability values range from zero to one, where one (highest possible reliability) means that all variation in the measure’s rates is the result of variation in differences in performance across agencies, while zero (lowest possible reliability) means that all variation is a result of measurement error.

Following the approach described by Adams,[[1]](#footnote-1) Acumen fit a beta-binomial model to estimate measure reliability. The beta-binomial model is appropriate because a particular agency’s measure rate follows a binomial distribution (i.e., all measures are pass/fail), and it is reasonable to assume that the agencies’ true measure rates vary and follow a beta distribution. The true measure rates among the agencies vary because of differences in agency styles, for example. It is reasonable to use the beta distribution to fit the true measure rates because it is a flexible distribution on the interval from 0 to 1, can have any mean on the interval, and can be skewed left, right, or U-shaped.

Equation (1), which is based on the beta-binomial model, shows that reliability is dependent on two variance components: the variation across agencies, and variation within agencies. In general, reliability for agencies will be higher when the measure rates across agencies are more heterogeneous (as measured by the agency-to-agency variation). Agencies with larger samples (n) and pass rates (p) nearer to 0 or 1 will have higher levels of reliability because the agency-specific error is reduced (i.e. the estimated agency rates are more precise).

(1)

Using the beta-binomial model, Acumen concluded that the measure reliability is high. Under the revised specifications, the 25th percentile reliability score is **0.76** (compared to a score of **0.51** when the measure is calculated using the current specifications), which is above the range considered acceptable for drawing inferences about home health agencies (i.e., 0.70 – 0.80). Table 4 summarizes the distribution of reliability scores for agencies with at least 20 valid episodes of care.

Table : Distribution of Beta Binomial Reliability Scores for Agencies with at Least 20 Valid Episodes

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Measure Specification** | **Mean** | **Min.** | **p10** | **p25** | **p50** | **p75** | **p90** | **Max** |
| Current Logic | 0.76 | 0.04 | 0.25 | 0.51 | 0.99 | 1.00 | 1.00 | 1.00 |
| Revised Logic | 0.84 | 0.21 | 0.59 | 0.76 | 0.90 | 0.97 | 1.00 | 1.00 |

## Intraclass Correlation Coefficient

Acumen also calculated the test-retest reliability using the ICC to measure between-agency variation and within-agency variation. Home health episodes within each agency were randomly divided into two separate equally-sized groups. Performance rates were obtained for each set within a measure. Then, a measure-level mean and variance were estimated using the paired performance rates, and an ICC statistic was derived. ICC values that approach 1 indicate that the fraction of the total variance due to between-agency variation is high. Under the revised specifications, the ICC is **0.78** for agencies with at least 40 valid episodes (compared to an ICC of **0.50** when the measure is calculated under the current specifications**)**, indicating that most of the total variation is due to between-agency variation.

# Measure Validity

Acumen evaluated the convergent validity of the revised measure. Convergent validity refers to the extent to which measures that are designed to assess the same construct are related to each other. To evaluate the convergent validity of the measure, Acumen calculated the Spearman rank correlations of the *Heart Failure Symptoms Addressed During All Episodes of Care* measure with other publicly-reported measures of home health quality derived from OASIS assessments; Table 5 shows the results. Correlation was low across all measures, showing that this measure captures a distinct type of care quality.

**Table 5: Spearman Rank Correlations of *Heart Failure Symptoms Addressed* with Other Measures of Home Health Quality**

|  |  |
| --- | --- |
| **Measure Performance** | **Spearman Rank Correlations** |
| Improvement in Ambulation | -0.1656 |
| Improvement in Bed Transfer | -0.1345 |
| Improvement in Bathing | -0.1122 |
| Improvement in Management of Oral Medications | -0.0989 |
| Improvement in Pain Interfering With Activity | -0.0088 |
| Improvement in Status of Surgical Wounds | 0.0527 |
| Emergency Department Use Without Hospitalization | -0.0066 |
| Acute Care Hospitalization | 0.0761 |
| Timely Initiation of Care | 0.0255 |
| Depression Assessment Conducted | 0.0867 |
| Multifactor Fall Risk Assessment Conducted | 0.1658 |
| Pressure Ulcer Risk Assessment Conducted | -0.0658 |
| Pressure Ulcer Treatment in Plan of Care | 0.0602 |
| Pressure Ulcer Prevention Implemented in All Episodes of Care | 0.0499 |
| Diabetic Foot Care Plan Implemented in All Episodes of Care | 0.0573 |
| Flu Immunization Received for Current Flu Season | 0.0152 |
| Pneumococcal Polysaccharide Vaccine Ever Received | -0.0211 |

# Exclusion Analysis

Acumen calculated the frequency of exclusions by exclusion type. Under the revised specifications, the measure excludes (i) home health episodes for which patient does not have a heart failure diagnosis, and (ii) home health episodes that end in death. Acumen found that the exclusions are supported by sufficient frequency of occurrence, and that the results would be distorted without the exclusions. Table 6 shows the number of quality episodes excluded as a result of each exclusion type.

Table : *Heart Failure Symptoms Addressed* Measure Exclusions, by Type

|  |  |
| --- | --- |
| **Exclusion Type** | **# of Quality Episodes Excluded** |
| Patient death at home (missing M2400A value) | 28,919 |
| No heart failure – no diagnosis found and symptoms not addressed | 2,936,811 |
| **Total # of Quality Episodes Excluded** | 2,965,730 |

Acumen also analyzed impact of excluding the long-term episodes. When NQF first considered the *Heart Failure Symptoms Addressed* measure in 2008, the NQF reviewers added a long-term episode exclusion to avoid excessive burden to agencies in reviewing records longer than 60 days. Therefore, the version of the *Heart Failure Symptoms Addressed* measure that was endorsed in 2011 included only the short-term episodes. However, the data needed to compute the measure are collected once at the end of each care episode, regardless of whether the episode is short or long-term, so including all episodes in the measure does not increase burden. In preparing this measure for the current NQF reevaluation process, Acumen re-considered the impact of the long-term episodes exclusion.

Removing the long-term episodes exclusion (i.e., using all episodes of care rather than the short-term episodes only) does *not* distort the results; the mean agency performance stays almost the same (~93.0%) as a result of this change. Additionally, removing the long-term episodes exclusion (i.e., using all episodes of care rather than the short-term episodes only) actually causes the number of agencies eligible for the measure to increase from 7,566 to 8,822 and the average number of valid episodes per agency to increase from 192 to 216. Table 7 shows the distribution of agency performance and number of agencies meeting the assessment threshold by episode type restriction.

**Table 7: Agency Performance and Number of Agencies Meeting Episode Threshold, by Episode Type Restriction**

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Episode Type** | **# of Agencies with ≥ 20 Episodes** | **Agency**  **Average** | **Std. Dev.** | **Average # of Episodes per Agency** | **Percentile** | | | | | | |
| **min** | **p10** | **p25** | **p50** | **p75** | **p90** | **max** |
| Short-Term Episodes Only | 7,566 | 92.8% | 7.3% | 192.1 | 0.0% | 84.6% | 90.3% | 94.4% | 97.4% | 100.0% | 100.0% |
| Long-Term Episodes Only | 5,221 | 93.2% | 7.0% | 77.0 | 1.5% | 84.8% | 90.7% | 95.0% | 97.7% | 100.0% | 100.0% |
| All-Episodes | 8,822 | 92.9% | 7.2% | 216.2 | 0.7% | 85.0% | 90.5% | 94.6% | 97.4% | 100.0% | 100.0% |

Table 8 - ICD-9 and ICD-10 Codes Defined as Heart Failure Diagnosis

| ICD-9 Codes | Corresponding ICD-10-CM Codes |
| --- | --- |
| **402.01** Hypertensive heart disease, malignant, with heart failure | **I11.0** Hypertensive heart disease with heart failure |
| **402.11** Hypertensive heart disease, benign, with heart failure |
| **402.91** Hypertensive heart disease, unspecified, with heart failure |
| **404.01** Hypertensive heart and chronic kidney disease, malignant, with heart failure and with chronic kidney disease stage I through stage IV, or unspecified | **I13.0** Hypertensive heart and chronic kidney disease with heart failure and stage 1 through stage 4 chronic kidney disease, or unspecified chronic kidney disease |
| **404.11** Hypertensive heart and chronic kidney disease, benign, with heart failure and with chronic kidney disease stage I through stage IV, or unspecified |
| **404.91** Hypertensive heart and chronic kidney disease, unspecified, with heart failure and with chronic kidney disease stage I through stage IV, or unspecified |
| **404.03** Hypertensive heart and chronic kidney disease, malignant, with heart failure and with chronic kidney disease stage V or end stage renal disease | **I13.2** Hypertensive heart and chronic kidney disease with heart failure and with stage 5 chronic kidney disease, or end stage renal disease |
| **404.13** Hypertensive heart and chronic kidney disease, benign, with heart failure and chronic kidney disease stage V or end stage renal disease |
| **404.93** Hypertensive heart and chronic kidney disease, unspecified, with heart failure and chronic kidney disease stage V or end stage renal disease |
| **428.0** Congestive heart failure, unspecified | **I50.9** Heart failure, unspecified |
| **428.1** Left heart failure | **I50.1** Left ventricular failure |
| **428.20** Unspecified systolic heart failure | **I50.20** Unspecified systolic (congestive) heart failure |
| **428.21** Acute systolic heart failure | **I50.21** Acute systolic (congestive) heart failure |
| **428.22** Chronic systolic heart failure | **I50.22** Chronic systolic (congestive) heart failure |
| **428.23** Acute on chronic systolic heart failure | **I50.23** Acute on chronic systolic (congestive) heart failure |
| **428.30** Unspecified diastolic heart failure | **I50.30** Unspecified diastolic (congestive) heart failure |
| **428.31** Acute diastolic heart failure | **I50.31** Acute diastolic (congestive) heart failure |
| **428.32** Chronic diastolic heart failure | **I50.32** Chronic diastolic (congestive) heart failure |
| **428.33** Acute on chronic diastolic heart failure | **I50.33** Acute on chronic diastolic (congestive) heart failure |
| **428.40** Unspecified combined systolic and diastolic heart failure | **I50.40** Unspecified combined systolic (congestive) and diastolic (congestive) heart failure |
| **428.41** Acute combined systolic and diastolic heart failure | **I50.41** Acute combined systolic (congestive) and diastolic (congestive) heart failure |
| **428.42** Chronic combined systolic and diastolic heart failure | **I50.42** Chronic combined systolic (congestive) and diastolic (congestive) heart failure |
| **428.43** Acute on chronic combined systolic and diastolic heart failure | **I50.43** Acute on chronic combined systolic (congestive) and diastolic (congestive) heart failure |
| **428.9** Heart failure, unspecified | **I50.9** Heart failure, unspecified |

1. For more information about reliability testing for performance measurement, as well as the methodology for constructing the reliability score reported on Table 6, see “Reliability of Provider Profiling: A Tutorial” by John Adams, RAND. http://www.rand.org/pubs/technical\_reports/TR653.html [↑](#footnote-ref-1)