

NATIONAL QUALITY FORUM

Measure Submission and Evaluation Worksheet 5.0

This form contains the information submitted by measure developers/stewards, organized according to NQF's measure evaluation criteria and process. The evaluation criteria, evaluation guidance documents, and a blank online submission form are available on the [submitting standards web page](#).

NQF #: 0506 NQF Project: Pulmonary Project
(for Endorsement Maintenance Review) Original Endorsement Date: Oct 28, 2008 Most Recent Endorsement Date: Oct 28, 2008 Last Updated Date: Oct 15, 2012
BRIEF MEASURE INFORMATION
De.1 Measure Title: Hospital 30-day, all-cause, risk-standardized readmission rate (RSRR) following pneumonia hospitalization
Co.1.1 Measure Steward: Centers for Medicare and Medicaid Services
De.2 Brief Description of Measure: The measure estimates a hospital-level risk-standardized readmission rate (RSRR) defined as unplanned readmissions for any cause within 30 days of the discharge date for the index hospitalization for patients discharged from the hospital with a principal diagnosis of pneumonia. The target population is patients 18 and over. CMS annually reports the measure for patients who are 65 years or older and are either enrolled in fee-for-service (FFS) Medicare and hospitalized in non-federal hospitals or are hospitalized in Veterans Health Administration (VA) facilities. Since NQF-endorsement, the measure has been tested and shown to perform well in an all-payer population aged 18 and older and has been re-specified for this broader age group. The full details of the all-payer analysis and testing are attached.
2a1.1 Numerator Statement: The outcome for this measure is 30 day all-cause readmission. We define all-cause readmission as an inpatient admission for any cause, with the exception of planned readmissions, within 30 days from the date of discharge from the index pneumonia admission. If a patient has one or more admissions (for any reason) within 30 days of the date of discharge of the index admission, only one was counted as a readmission. For the detailed definition of planned readmissions, please refer to the attached report, Respecifying the Hospital 30-Day Pneumonia and 30-Day Chronic Obstructive Pulmonary Disease Readmission Measures by adding a Planned Readmission Algorithm. The numerator of the risk-adjusted ratio is the predicted number of readmissions within 30 days given the hospital's performance with its observed case mix. The term "predicted" describes the numerator result, which is calculated using the hospital-specific intercept term. (See details below in the 2a1.13 Statistical risk model and variables.)
2a1.4 Denominator Statement: The cohort includes admissions for patients 18 and over hospitalized for pneumonia. The measure is currently publicly reported by CMS for patients 65 years and older who are either enrolled in Medicare FFS and admitted to non-federal hospitals, or admitted to VA hospitals. The measure includes admissions for patients discharged from the hospital with a principal diagnosis of pneumonia and with a complete claims history for the 12 months prior to admission.
2a1.8 Denominator Exclusions: The measure excludes admissions for patients: For all cohorts, the measure excludes admissions for patients: <ul style="list-style-type: none">• with an in-hospital death (because they are not eligible for readmission);• transferred to another acute care hospital (because the readmission is attributed to the hospital that discharges the patient to a non-acute setting);• discharged against medical advice (AMA) (because providers did not have the opportunity to deliver full care and prepare the patient for discharge);• admitted with pneumonia within 30 days of discharge from a qualifying index admission (Admissions within 30 days of

discharge of an index admission will be considered readmissions. No admission is counted as a readmission and an index admission. The next eligible admission after the 30-day time period following an index admission will be considered another index admission.)

For Medicare FFS patients, the measure additionally excludes admissions for patients:

- without at least 30 days post-discharge enrollment in FFS Medicare (because the 30-day readmission outcome cannot be assessed in this group).

1.1 Measure Type: Outcome

2a1. 25-26 Data Source: Administrative claims

2a1.33 Level of Analysis: Facility

1.2-1.4 Is this measure paired with another measure? No

De.3 If included in a composite, please identify the composite measure (title and NQF number if endorsed):

This measure is not formally paired with another measure, however this measure is harmonized with a measure of hospital-level, all-cause, 30-day, risk-standardized mortality following a pneumonia hospitalization.

STAFF NOTES (issues or questions regarding any criteria)

Comments on Conditions for Consideration:

Is the measure untested? Yes No If untested, explain how it meets criteria for consideration for time-limited endorsement:

1a. Specific national health goal/priority identified by DHHS or NPP addressed by the measure (check De.5):

5. Similar/related endorsed or submitted measures (check 5.1):

Other Criteria:

Staff Reviewer Name(s):

1. IMPACT, OPPORTUNITY, EVIDENCE - IMPORTANCE TO MEASURE AND REPORT

Importance to Measure and Report is a threshold criterion that must be met in order to recommend a measure for endorsement. All three subcriteria must be met to pass this criterion. See [guidance on evidence](#).

Measures must be judged to be important to measure and report in order to be evaluated against the remaining criteria. (evaluation criteria)

1a. High Impact: H M L I

(The measure directly addresses a specific national health goal/priority identified by DHHS or NPP, or some other high impact aspect of healthcare.)

De.4 Subject/Topic Areas (Check all the areas that apply): Pulmonary/Critical Care : Pneumonia

De.5 Cross Cutting Areas (Check all the areas that apply): Care Coordination, Overuse, Population Health, Safety : Complications, Safety : Healthcare Associated Infections

1a.1 Demonstrated High Impact Aspect of Healthcare: Affects large numbers, A leading cause of morbidity/mortality, High resource use, Patient/societal consequences of poor quality, Severity of illness

1a.2 If "Other," please describe:

1a.3 Summary of Evidence of High Impact (Provide epidemiologic or resource use data):

The Medicare Payment Advisory Commission (MedPAC) has called for hospital-specific public reporting of readmission rates, identifying pneumonia as a priority condition (MedPAC, 2007). MedPAC finds that readmissions are common, costly, and often preventable. Based on 2005 Medicare data, MedPAC estimates that about 8.9% of Medicare pneumonia admissions were followed

by a readmission within 15 days, accounting for more than 74,000 admissions at a cost of \$533 million.

Pneumonia results in approximately 1.2 million hospital admissions each year and accounts for more than \$10 billion annually in hospital expenditures. Among patients over 65 years of age, it is the second leading cause of hospitalization, and is the leading infectious cause of death (Lindenauer et al., 2011). Approximately 20% of pneumonia patients were rehospitalized within thirty days, representing the second-highest proportion of all rehospitalizations at 6.3% (Jencks 2009). Pneumonia readmission is a costly event and represents an undesirable outcome of care from the patient's perspective, and highly disparate pneumonia readmission rates among hospitals suggest there is room for improvement. (MedPAC 2007, Bernheim 2010)

1a.4 Citations for Evidence of High Impact cited in 1a.3: Bernheim SM, et al. 2010 Measures Maintenance Technical Report: Acute Myocardial Infarction, Heart Failure and Pneumonia 30-day Risk Standardized Mortality Rate. 2010 Available at: <http://www.qualitynet.org/dcs/ContentServer?c=Page&pagename=QnetPublic/Page/QnetTier3&cid=1163010421830>

Jencks SF, Williams MV, Coleman EA. Rehospitalizations among patients in the Medicare fee-for-service program. N Engl J Med. 2009 Apr 2;360(14):1418-28.

Lindenauer PK, Normand SL, Drye EE, et al. Development, validation, and results of a measure of 30-day readmission following hospitalization for pneumonia. J Hosp Med. 2011;6(3):142-150

Report to the Congress: Promoting Greater Efficiency in Medicare. Washington, DC: Medicare Payment Advisory Commission, 2007.

1b. Opportunity for Improvement: H M L I

(There is a demonstrated performance gap - variability or overall less than optimal performance)

1b.1 Briefly explain the benefits (improvements in quality) envisioned by use of this measure:

The goal of this measure is to improve patient outcomes by providing patients, physicians, and hospitals with information about hospital-level, risk-standardized readmission rates following hospitalization for pneumonia. Measurement of patient outcomes allows for a broad view of quality of care that encompasses more than what can be captured by individual process-of-care measures. Complex and critical aspects of care, such as communication between providers, prevention of, and response to, complications, patient safety and coordinated transitions to the outpatient environment, all contribute to patient outcomes but are difficult to measure by individual process measures. The goal of outcomes measurement is to risk-adjust for patients' conditions at the time of hospital admission and then evaluate patient outcomes. This readmission measure was developed to identify institutions, whose performance is better or worse than would be expected based on their patient case-mix, and therefore promote hospital quality improvement and better inform consumers about care quality.

1b.2 Summary of Data Demonstrating Performance Gap (Variation or overall less than optimal performance across providers): **[For Maintenance** – Descriptive statistics for performance results for this measure - distribution of scores for measured entities by quartile/decile, mean, median, SD, min, max, etc.]

There is substantial variation in pneumonia RSRRs among hospitals. Using Medicare FFS data from July 2008 to June 2011 and the updated measure (with the new planned readmission algorithm), the median hospital RSRR for pneumonia was 17.7% with a range of 13.4% to 25.5%. The 5th percentile was 15.7% and the 95th percentile was 20.5%. The interquartile range was 16.9% to 18.7%.

1b.3 Citations for Data on Performance Gap: **[For Maintenance** – Description of the data or sample for measure results reported in 1b.2 including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities included]

Lindenauer PK, Bernheim SM, Grady JN, et al. The performance of US hospitals as reflected in risk-standardized 30-day mortality and readmission rates for medicare beneficiaries with pneumonia. J Hosp Med. 2010;5(6):E12-E18

1b.4 Summary of Data on Disparities by Population Group: **[For Maintenance** – Descriptive statistics for performance results for this measure by population group]

The measure is a hospital-level measure and therefore CMS assessed evidence of disparities by examining hospital performance

based on the proportion of African-American patients served or the proportion of low-income patients served by the hospital.

The analyses examining the proportion of African-American patients that a hospital served show slightly higher RSRRs for hospitals with higher proportions of African-American patients compared with lower proportions, but the range of performance across all levels is similar. We divided hospitals into deciles based on the proportion of their patients that were African-American and looked at hospital performance on the measure across deciles. The combined lowest 5 deciles include hospitals with fewer than 5% African-American patients and have a median pneumonia RSRR of 17.8% (range 12.7% -27.4%) in comparison hospitals in the highest decile with greater than 25% African American patients have a median pneumonia RSRR of 19.2% (range 15.7% - 27.0%). Although this demonstrates slightly worse performance of hospitals with a large proportion of African-American patients, these analyses also show wide variation in performance of hospitals regardless of the proportion of African-American patients and suggest that hospitals with large proportions of African American patients are not consistently performing at a lower or higher level than other hospitals.

Similar analyses were completed to evaluate hospital differences in performance on RSRR based on the socioeconomic status of their patients. These analyses suggest a slightly higher median RSRR at the hospitals in the lowest quartile based on the socioeconomic status of their patients (as measured by the median of the patient's zip-code level median income). The lowest quartile hospitals have a median RSRR of 18.4% compared to a median RSRR of 18.1% for hospitals in highest quartile. However, the range for the two groups is largely overlapping (14.2% - 27.4 % vs. 18.1% - 25.6%) demonstrating that substantial numbers of hospitals serving low SES patients perform well on the measure.

We conducted these analyses prior to adding the planned readmission algorithm, but do not expect the change would substantively affect the results.

Overall these analyses provide little compelling evidence of clinically significant disparities at the hospital level.

1b.5 Citations for Data on Disparities Cited in 1b.4: [For Maintenance – Description of the data or sample for measure results reported in 1b.4 including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities included]

The sample for the above analyses is from a similar 3 year cohort of hospitalizations as the data for the performance gap analysis above (January 2007- December 2009) but limited to hospitals with at least 25 pneumonia cases over the 3 year period, a total of 4,925 hospitals (without the new planned readmission algorithm).

Medicare Hospital Quality Chartbook 2010. Performance Report on Outcomes Measures for Acute Myocardial Infarction, Heart Failure, and Pneumonia. Yale New Haven Health System Corporation/Center for Outcomes Research and Evaluation. New Haven, CT: 2010. (http://cms.gov/HospitalQualityInits/20_OutcomeMeasures.asp#TopOfPage)

1c. Evidence (Measure focus is a health outcome OR meets the criteria for quantity, quality, consistency of the body of evidence.)
Is the measure focus a health outcome? Yes No If not a health outcome, rate the body of evidence.

Quantity: H M L I **Quality:** H M L I **Consistency:** H M L I

Quantity	Quality	Consistency	Does the measure pass subcriterion1c?
M-H	M-H	M-H	Yes <input type="checkbox"/>
L	M-H	M	Yes <input type="checkbox"/> IF additional research unlikely to change conclusion that benefits to patients outweigh harms: otherwise No <input type="checkbox"/>
M-H	L	M-H	Yes <input type="checkbox"/> IF potential benefits to patients clearly outweigh potential harms: otherwise No <input type="checkbox"/>
L-M-H	L-M-H	L	No <input type="checkbox"/>

Health outcome – rationale supports relationship to at least one healthcare structure, process, intervention, or service **Does the measure pass subcriterion1c?**
Yes IF rationale supports relationship

1c.1 Structure-Process-Outcome Relationship (Briefly state the measure focus, e.g., health outcome, intermediate clinical

outcome, process, structure; then identify the appropriate links, e.g., structure-process-health outcome; process- health outcome; intermediate clinical outcome-health outcome):

This measure calculates hospital-level, 30-day all-cause readmission rates after hospitalization for pneumonia. The goal is to directly affect patient outcomes by measuring risk-standardized rates of readmission.

1c.2-3 Type of Evidence (Check all that apply):

Other

N/A This is an outcomes measure, not a process measure.

1c.4 Directness of Evidence to the Specified Measure (State the central topic, population, and outcomes addressed in the body of evidence and identify any differences from the measure focus and measure target population):

N/A This is an outcomes measure, not a process measure.

1c.5 Quantity of Studies in the Body of Evidence (Total number of studies, not articles): N/A This is an outcomes measure, not a process measure.

1c.6 Quality of Body of Evidence (Summarize the certainty or confidence in the estimates of benefits and harms to patients across studies in the body of evidence resulting from study factors. Please address: a) study design/flaws; b) directness/indirectness of the evidence to this measure (e.g., interventions, comparisons, outcomes assessed, population included in the evidence); and c) imprecision/wide confidence intervals due to few patients or events): N/A This is an outcomes measure, not a process measure.

1c.7 Consistency of Results across Studies (Summarize the consistency of the magnitude and direction of the effect): N/A This is an outcomes measure, not a process measure.

1c.8 Net Benefit (Provide estimates of effect for benefit/outcome; identify harms addressed and estimates of effect; and net benefit - benefit over harms):

N/A This is an outcomes measure, not a process measure

1c.9 Grading of Strength/Quality of the Body of Evidence. Has the body of evidence been graded? No

1c.10 If body of evidence graded, identify the entity that graded the evidence including balance of representation and any disclosures regarding bias: N/A This is an outcomes measure, not a process measure.

1c.11 System Used for Grading the Body of Evidence: Other

1c.12 If other, identify and describe the grading scale with definitions: N/A This is an outcomes measure, not a process measure

1c.13 Grade Assigned to the Body of Evidence: N/A This is an outcomes measure, not a process measure.

1c.14 Summary of Controversy/Contradictory Evidence: All-cause Readmission

This measure calculates a 30-day all-cause readmission rate. CMS measures all-cause readmission rather than readmission due to certain conditions (e.g. heart failure readmissions) for a number of reasons. First, a narrow focus on specific causes of readmission may simply provide an incentive to shift patients away from those codes. Second, within the chain of events that lead to a patient being readmitted to the hospital there is often some aspect of care that could be improved, thereby reducing the risk of readmission. This is not to suggest that all readmissions are preventable, but the goal of the measure is to encourage broad approaches to quality improvement which will thereby lower all patients' risk of readmission. More narrowly defining readmission measures to those that are disease specific may incentivize a limited focus on improvements in care as opposed to thinking comprehensively about the patient's full medical and social needs at discharge. Factors which may influence readmission rates include medication reconciliation, patient education, follow-up care and communication between inpatient and outpatient providers. The goal is not to

reduce the readmission rate to zero but to reduce overall readmission rates to what is achievable by the best hospitals.

1c.15 Citations for Evidence other than Guidelines(Guidelines addressed below):

N/A This is an outcomes measure, not a process measure.

1c.16 Quote verbatim, the specific guideline recommendation (Including guideline # and/or page #):

N/A This is an outcomes measure, not a process measure.

1c.17 Clinical Practice Guideline Citation: N/A This is an outcomes measure, not a process measure.

1c.18 National Guideline Clearinghouse or other URL: N/A This is an outcomes measure, not a process measure.

1c.19 Grading of Strength of Guideline Recommendation. Has the recommendation been graded? No

1c.20 If guideline recommendation graded, identify the entity that graded the evidence including balance of representation and any disclosures regarding bias:

1c.21 System Used for Grading the Strength of Guideline Recommendation: Other

1c.22 If other, identify and describe the grading scale with definitions: N/A This is an outcomes measure, not a process measure.

1c.23 Grade Assigned to the Recommendation: N/A This is an outcomes measure, not a process measure.

1c.24 Rationale for Using this Guideline Over Others: N/A This is an outcomes measure, not a process measure.

Based on the NQF descriptions for rating the evidence, what was the developer's assessment of the quantity, quality, and consistency of the body of evidence?

1c.25 Quantity: High 1c.26 Quality: High 1c.27 Consistency: High

1c.28 Attach evidence submission form:

1c.29 Attach appendix for supplemental materials:

Was the threshold criterion, *Importance to Measure and Report*, met?

(1a & 1b must be rated moderate or high and 1c yes) Yes No

Provide rationale based on specific subcriteria:

For a new measure if the Committee votes NO, then STOP.

For a measure undergoing endorsement maintenance, if the Committee votes NO because of 1b. (no opportunity for improvement), it may be considered for continued endorsement and all criteria need to be evaluated.

2. RELIABILITY & VALIDITY - SCIENTIFIC ACCEPTABILITY OF MEASURE PROPERTIES

Extent to which the measure, as specified, produces consistent (reliable) and credible (valid) results about the quality of care when implemented. (**evaluation criteria**)

Measure testing must demonstrate adequate reliability and validity in order to be recommended for endorsement. Testing may be conducted for data elements and/or the computed measure score. Testing information and results should be entered in the appropriate field. Supplemental materials may be referenced or attached in item 2.1. See [guidance on measure testing](#).

S.1 Measure Web Page (In the future, NQF will require measure stewards to provide a URL link to a web page where current detailed specifications can be obtained). Do you have a web page where current detailed specifications for this measure can be obtained? Yes

S.2 If yes, provide web page URL: www.qualitynet.org

2a. RELIABILITY. Precise Specifications and Reliability Testing: H M L I

2a1. Precise Measure Specifications. (*The measure specifications precise and unambiguous.*)

2a1.1 Numerator Statement (*Brief, narrative description of the measure focus or what is being measured about the target population, e.g., cases from the target population with the target process, condition, event, or outcome*):

The outcome for this measure is 30 day all-cause readmission. We define all-cause readmission as an inpatient admission for any cause, with the exception of planned readmissions, within 30 days from the date of discharge from the index pneumonia admission. If a patient has one or more admissions (for any reason) within 30 days of the date of discharge of the index admission, only one was counted as a readmission. For the detailed definition of planned readmissions, please refer to the attached report, Respecifying the Hospital 30-Day Pneumonia and 30-Day Chronic Obstructive Pulmonary Disease Readmission Measures by adding a Planned Readmission Algorithm.

The numerator of the risk-adjusted ratio is the predicted number of readmissions within 30 days given the hospital's performance with its observed case mix. The term "predicted" describes the numerator result, which is calculated using the hospital-specific intercept term. (See details below in the 2a1.13 Statistical risk model and variables.)

2a1.2 Numerator Time Window (*The time period in which the target process, condition, event, or outcome is eligible for inclusion*): We define this as readmission for any cause within 30 days from the date of discharge of the index pneumonia hospitalization.

2a1.3 Numerator Details (*All information required to identify and calculate the cases from the target population with the target process, condition, event, or outcome such as definitions, codes with descriptors, and/or specific data collection items/responses*): Note: This outcome measure does not have a traditional numerator and denominator like a core process measure (e.g., percentage of adult patients with diabetes aged 18-75 years receiving one or more hemoglobin A1c tests per year); thus, we use this field to define the measure outcome.

The measure counts readmissions to any acute care hospital for any cause within 30 days of the date of discharge of the index pneumonia admission.

Planned admissions not counted as readmissions

Unplanned readmissions are acute clinical events experienced by a patient that require urgent hospitalizations. Higher than expected unplanned readmission rates suggest lower quality of hospital and post-discharge care and are the focus of hospital quality measurement as part of quality improvement efforts. In contrast, planned readmissions are generally not a signal of quality of care. Furthermore, there is concern that including planned readmissions in a readmission measure could create a disincentive to provide appropriate care to patients who are scheduled for elective or necessary procedures, unrelated to the quality of the prior admission, within 30 days of discharge. We have, therefore, developed an algorithm for using claims data to identify "planned readmissions" that will not count as outcomes in the readmission measure.

In Medicare FFS data from the July 2008 to June 2011, 0.6% of index hospitalizations for pneumonia were followed by a planned readmission within 30 days of discharge. After accounting for planned readmissions, the crude 30-day measure readmission rate decreased from 18.5% to 17.8%.

The detailed algorithm for identifying planned readmissions is in the attached report, Respecifying the Hospital 30-Day Pneumonia and 30-Day Chronic Obstructive Pulmonary Disease Readmission Measures by adding a Planned Readmission Algorithm.

2a1.4 Denominator Statement (*Brief, narrative description of the target population being measured*):

The cohort includes admissions for patients 18 and over hospitalized for pneumonia. The measure is currently publicly reported by CMS for patients 65 years and older who are either enrolled in Medicare FFS and admitted to non-federal hospitals, or admitted to VA hospitals.

The measure includes admissions for patients discharged from the hospital with a principal diagnosis of pneumonia and with a complete claims history for the 12 months prior to admission.

2a1.5 Target Population Category (Check all the populations for which the measure is specified and tested if any): [Adult/Elderly Care, Populations at Risk](#)

2a1.6 Denominator Time Window (The time period in which cases are eligible for inclusion):

This measure was developed with 12 months of data. Currently the measure is publicly-reported with three years of index hospitalizations.

2a1.7 Denominator Details (All information required to identify and calculate the target population/denominator such as definitions, codes with descriptors, and/or specific data collection items/responses):

This outcome measure does not have a traditional numerator and denominator like a core process measure (e.g., percentage of adult patients with diabetes aged 18-75 years receiving one or more hemoglobin A1c tests per year); thus, we use this field to define the measure cohort.

The denominator includes patients 18 and over hospitalized for pneumonia. The measure is currently publicly reported by CMS for patients 65 years and older who are either enrolled in Medicare FFS and admitted to non-federal hospitals, or admitted to a VA hospital. To be included in the Medicare FFS cohort the patients must have been continuously enrolled in Medicare FFS Parts A and B for the 12 months prior to the index hospitalization.

The denominator includes admissions for patients discharged from the hospital with a principal diagnosis of pneumonia (ICD-9-CM codes 480.0, 480.1, 480.2, 480.3, 480.8, 480.9, 481, 482.0, 482.1, 482.2, 482.30, 482.31, 482.32, 482.39, 482.40, 482.41, 482.42, 482.49, 482.81, 482.82, 482.83, 482.84, 482.89, 482.9, 483.0, 483.1, 483.8, 485, 486, 487.0, and 488.11; ICD-10-CM codes J120, J121, J122, J1281, J1289, J129, J13, J181, J150, J151, J14, J154, J154, J153, J154, J1520, J1521, J1521, Z16, J1529, J158, J155, J156, A481, J158, J159, J157, J160, J168, J180, J189, J1100, J129, J09119).

2a1.8 Denominator Exclusions (Brief narrative description of exclusions from the target population):

The measure excludes admissions for patients:

For all cohorts, the measure excludes admissions for patients:

- with an in-hospital death (because they are not eligible for readmission);
- transferred to another acute care hospital (because the readmission is attributed to the hospital that discharges the patient to a non-acute setting);
- discharged against medical advice (AMA) (because providers did not have the opportunity to deliver full care and prepare the patient for discharge);
- admitted with pneumonia within 30 days of discharge from a qualifying index admission (Admissions within 30 days of discharge of an index admission will be considered readmissions. No admission is counted as a readmission and an index admission. The next eligible admission after the 30-day time period following an index admission will be considered another index admission.)

For Medicare FFS patients, the measure additionally excludes admissions for patients:

- without at least 30 days post-discharge enrollment in FFS Medicare (because the 30-day readmission outcome cannot be assessed in this group).

2a1.9 Denominator Exclusion Details (All information required to identify and calculate exclusions from the denominator such as definitions, codes with descriptors, and/or specific data collection items/responses):

Measure exclusions are determined as follows

For all cohorts, the measure excludes admissions for patients:

- Admissions with an in-hospital death are identified in the discharge disposition indicator in claims data.
- Admissions for patients who were transferred to another acute care hospital or VA hospital are identified in the claims when a patient with a qualifying admission is discharged from an acute care hospital and admitted to another acute care hospital on the same day or next day;

- Discharges against medical advice (AMA) are identified by examining the discharge destination indicator in claims data;
- Pneumonia admissions within 30 days of discharge from a qualifying index admission are identified by comparing the discharge date from the index admission with the readmission date

For Medicare FFS patients, the measure additionally excludes admissions for patients who:

- Admissions without at least 30 days post-discharge enrollment in FFS Medicare is obtained by examining the Medicare Enrollment Database (EDB)

2a1.10 Stratification Details/Variables (All information required to stratify the measure results including the stratification variables, codes with descriptors, definitions, and/or specific data collection items/responses):

N/A

2a1.11 Risk Adjustment Type (Select type. Provide specifications for risk stratification in 2a1.10 and for statistical model in 2a1.13): **Statistical risk model** **2a1.12 If "Other," please describe:**

2a1.13 Statistical Risk Model and Variables (Name the statistical method - e.g., logistic regression and list all the risk factor variables. Note - risk model development should be addressed in 2b4.):

Our approach to risk adjustment is tailored to and appropriate for a publicly reported outcome measure, as articulated in the American Heart Association (AHA) Scientific Statement, "Standards for Statistical Models Used for Public Reporting of Health Outcomes" (Krumholz et. al., 2006).

The proposed measure employs a hierarchical logistic regression model to create a hospital level 30-day RSRR. In brief, the approach simultaneously models two levels (patient and hospital) to account for the variance in patient outcomes within and between hospitals (Normand & Shahian, 2007). At the patient level, each model adjusts the log-odds of readmission within 30-days of discharge for age and selected clinical covariates. The second level models the hospital-specific intercepts as arising from a normal distribution. The hospital intercept represents the underlying risk of readmission, after accounting for patient risk. See section 2a1.20. Calculation Algorithm/Measure Logic for more detail.

Candidate and Final Risk-adjustment Variables: Candidate variables were patient-level risk-adjustors that were expected to be predictive of readmission, based on empirical analysis, prior literature, and clinical judgment, including age and indicators of comorbidity and disease severity. For each patient, covariates are obtained from Medicare claims extending 12 months prior to and including the index admission. The model adjusts for case mix differences based on the clinical status of patients at the time of admission. We use condition categories (CCs), which are clinically meaningful groupings of more than 15,000 ICD-9-CM diagnosis codes. A file which contains a list of the ICD-9-CM codes and their groupings into CCs is available at <http://www.qualitynet.org/dcs/ContentServer?c=Page&pagename=QnetPublic%2FPage%2FQnetTier3&cid=1182785083979>. In addition, only comorbidities that convey information about the patient at admission or in the 12-months prior, and not complications that arise during the course of the hospitalization, are included in the risk-adjustment. Hence, we do not risk adjust for CCs that may represent adverse events of care and that are only recorded in the index admission.

The final set of risk-adjustment variables is:

Demographics

Age-65 (years above 65, continuous)

Male

Comorbidities

History of coronary artery bypass graft (CABG) surgery

History of infection (CC 1, 3-6)

Septicemia/shock (CC 2)

Metastatic cancer and acute leukemia (CC7)

Lung, upper digestive tract, and other severe cancers (CC8)

Lymphatic, head and neck, brain, and other major cancers; breast, prostate, colorectal and other cancers and tumors (CC 9-10)

See Guidance for Definitions of Rating Scale: H=High; M=Moderate; L=Low; I=Insufficient; NA=Not Applicable

Created on: 10/22/2012 at 10:44 AM

Diabetes mellitus (DM) and DM complications (CC 15-20, 119-120)
Protein-calorie malnutrition (CC 21)
Disorders of fluid/electrolyte/acid-base (CC 22-23)
Other gastrointestinal disorders (CC 36)
Severe hematological disorders (CC 44)
Iron deficiency and other/unspecified anemias and blood disease (CC 47)
Dementia and senility (CC 49-50)
Drug/alcohol abuse/dependence/psychosis (CC 51-53)
Major psychiatric disorders (CC 54-56)
Other psychiatric disorders (CC 60)
Hemiplegia, paraplegia, paralysis, functional disability (CC67-69, 100-102, 177-178)
Cardio-respiratory failure and shock (CC 79)
Congestive heart failure (CC 80)
Acute coronary syndrome (CC 81-82)
Chronic atherosclerosis (CC 83-84)
Valvular and rheumatic heart disease (CC 86)
Arrhythmias (CC 92-93)
Stroke (CC 95-96)
Vascular or circulatory disease (CC 104-106)
Chronic obstructive pulmonary disease (CC 108)
Fibrosis of lung and other chronic lung disorders (CC 109)
Asthma (CC 110)
Pneumonia (CC 111-113)
Pleural effusion/pneumothorax (CC 114)
Other lung disorders (CC 115)
End-stage renal disease or dialysis (CC 129-130)
Renal failure (CC 131)
Urinary tract infection (CC 135)
Other urinary tract disorders (CC 136)
Decubitus ulcer or chronic skin ulcer (CC 148-149)
Vertebral fractures (CC 157)
Other injuries (CC 162)

References:

Krumholz HM, Brindis RG, Brush JE, et al. 2006. Standards for Statistical Models Used for Public Reporting of Health Outcomes: An American Heart Association Scientific Statement From the Quality of Care and Outcomes Research Interdisciplinary Writing Group: Cosponsored by the Council on Epidemiology and Prevention and the Stroke Council Endorsed by the American College of Cardiology Foundation. *Circulation* 113: 456-462.

Normand S-LT, Shahian DM. 2007. Statistical and Clinical Aspects of Hospital Outcomes Profiling. *Stat Sci* 22 (2): 206-226.

2a1.14-16 Detailed Risk Model Available at Web page URL (or attachment). Include coefficients, equations, codes with descriptors, definitions, and/or specific data collection items/responses. Attach documents only if they are not available on a webpage and keep attached file to 5 MB or less. NQF strongly prefers you make documents available at a Web page URL. Please supply login/password if needed:

URL

<http://www.qualitynet.org/dcs/ContentServer?c=Page&pagename=QnetPublic%2FPage%2FQnetTier4&cid=1219069855841>

N/A

2a1.17-18. Type of Score: [Rate/proportion](#)

2a1.19 Interpretation of Score (Classifies interpretation of score according to whether better quality is associated with a higher score, a lower score, a score falling within a defined interval, or a passing score): **Better quality = Lower score**

2a1.20 Calculation Algorithm/Measure Logic(Describe the calculation of the measure score as an ordered sequence of steps including identifying the target population; exclusions; cases meeting the target process, condition, event, or outcome; aggregating data; risk adjustment; etc.):

The proposed measure employs a hierarchical logistic regression model to create a hospital level 30-day RSRR. In brief, the approach simultaneously models two levels (patient and hospital) to account for the variance in patient outcomes within and between hospitals (Normand & Shahian, 2007). At the patient level, each model adjusts the log-odds of readmission within 30-days of discharge for age and selected clinical covariates. The second level models the hospital-specific intercepts as arising from a normal distribution. The hospital intercept represents the underlying risk of readmission, after accounting for patient risk. The hospital-specific intercepts are given a distribution in order to account for the clustering (non-independence) of patients within the same hospital. If there were no differences among hospitals, then after adjusting for patient risk, the hospital intercepts should be identical across all hospitals.

The RSRR is calculated as the ratio of the number of “predicted” to the number of “expected” readmissions, multiplied by the national unadjusted readmission rate. For each hospital, the numerator of the ratio (“predicted”) is the number of readmissions within 30 days predicted on the basis of the hospital’s performance with its observed case mix, and the denominator (“expected”) is the number of readmissions expected on the basis of the nation’s performance with that hospital’s case mix. This approach is analogous to a ratio of “observed” to “expected” used in other types of statistical analyses. It conceptually allows for a comparison of a particular hospital’s performance given its case-mix to an average hospital’s performance with the same case-mix. Thus, a lower ratio indicates lower-than-expected readmission or better quality and a higher ratio indicates higher-than-expected readmission or worse quality.

The predicted hospital outcome (the numerator) is the sum of predicted probabilities of readmissions for all patients at a particular hospital. The predicted probability of each patient in that hospital is calculated using the hospital-specific intercept and patient risk factors. The expected number of readmissions (the denominator) is the sum of expected probabilities of readmission for all patients at a hospital. The expected probability of each patient in a hospital is calculated using a common intercept and patient risk factors.

References:

Normand S-LT, Shahian DM. 2007. Statistical and Clinical Aspects of Hospital Outcomes Profiling. Stat Sci 22 (2): 206-226.

2a1.21-23 Calculation Algorithm/Measure Logic Diagram URL or attachment:

URL

<http://www.qualitynet.org/dcs/ContentServer?c=Page&pagename=QnetPublic%2FPage%2FQnetTier4&cid=1219069855841>

2a1.24 Sampling (Survey) Methodology. If measure is based on a sample (or survey), provide instructions for obtaining the sample, conducting the survey and guidance on minimum sample size (response rate):

N/A –This measure is not based on a sample or survey.

2a1.25 Data Source (Check all the sources for which the measure is specified and tested). If other, please describe:

Administrative claims

2a1.26 Data Source/Data Collection Instrument (Identify the specific data source/data collection instrument, e.g. name of database, clinical registry, collection instrument, etc.): **Data sources for the FFS measure:**

1. Medicare Part A inpatient and Part B outpatient claims: This database contains claims data for fee-for service inpatient and outpatient services including: Medicare inpatient hospital care, outpatient hospital services, skilled nursing facility care, some home health agency services, as well as inpatient and outpatient claims for the 12 months prior to an index admission.

2. Medicare Enrollment Database (EDB): This database contains Medicare beneficiary demographic, benefit/coverage, and vital status information. This dataset was used to obtain information on several inclusion/exclusion indicators such as Medicare status on

admission as well as vital status. These data have previously been shown to accurately reflect patient vital status (Fleming Fisher et al., 1992).

The measure was originally developed with claims data from a 2006 sample of 226,545 cases from 4,675 hospitals. We have maintained and re-evaluated the models each year since public reporting of the measure began in 2009.

Fleming C., Fisher ES, Chang CH, Bubolz D, Malenda J. Studying outcomes and hospital utilization in the elderly: The advantages of a merged data base for Medicare and Veterans Affairs Hospitals. *Medical Care*. 1992; 30(5): 377-91.

Data sources for the all-payer update

For our analyses, we used all-payer data from California in addition to CMS data for Medicare FFS 65+ patients in California hospitals. California is a diverse state, and, with more than 37 million residents, California represents 12% of the US population. We used the California Patient Discharge Data, a large, linked database of patient hospital admissions. In 2006, there were approximately 3 million adult discharges from more than 450 non-Federal acute care hospitals. Records are linked by a unique patient identification number, allowing us to determine patient history from previous hospitalizations and to evaluate rates of both readmission and mortality (via linking with California vital statistics records).

Using all-payer data from California as well as CMS Medicare FFS data for California hospitals, we performed analyses to determine whether the publicly reported measures can be applied to all adult patients, including not only FFS Medicare patients aged 65+ but also non-FFS Medicare patients aged 65+ and younger patients aged 18-64 years at the time of admission.

2a1.27-29 Data Source/data Collection Instrument Reference Web Page URL or Attachment: [Attachment](#)
508 compliant pneumonia ICD-10 map-634623950487720270.pdf

2a1.30-32 Data Dictionary/Code Table Web Page URL or Attachment:

[URL](#)

Condition Category/ICD-9 Code Map available at:

(<http://www.qualitynet.org/dcs/ContentServer?c=Page&pagename=QnetPublic%2FPage%2FQnetTier3&cid=1182785083979>) See attached ICD-9 to ICD-10 crosswalk

2a1.33 Level of Analysis (Check the levels of analysis for which the measure is specified and tested): [Facility](#)

2a1.34-35 Care Setting (Check all the settings for which the measure is specified and tested): [Hospital/Acute Care Facility](#)

2a2. Reliability Testing. (Reliability testing was conducted with appropriate method, scope, and adequate demonstration of reliability.)

2a2.1 Data/Sample (Description of the data or sample including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities included):

Medicare Part A inpatient and outpatient claims data for calendar years 2007-2009 were used to test reliability (without including the new planned readmission algorithm). Specifically, two datasets were used to assess reliability:

2007-2009 Subset A1 (599,723 index hospital stays at 4,896 hospitals)

2007-2009 Subset A2 (599,601 hospital stays at 4,896 hospitals)

(Of note, Subsets A1 and A2 were created by randomly splitting the full 2007-2009 dataset)

2a2.2 Analytic Method (Describe method of reliability testing & rationale):

[Data element reliability](#)

In constructing the measures we aim to utilize only those data elements from the claims that have both face validity and reliability.

We avoid the use of fields that are thought to be coded inconsistently across hospitals or providers. Specifically, we use fields that are consequential for payment and which are audited. We identify such variables through empiric analyses and our understanding of CMS auditing and billing policies and seek to avoid variables which do not meet this standard. For example, "discharge disposition" is a variable in Medicare claims data that is not thought to be a reliable variable for identifying a transfer between two acute care facilities. Thus, we derive a variable using admission and discharge dates as a surrogate for "discharge disposition" to identify hospital admissions involving transfers. This allows us to identify these admissions using variables in the claims data which have greater reliability than the "discharge disposition" variable.

In addition, CMS has in place several hospital auditing programs used to assess overall claims code accuracy, to ensure appropriate billing, and for overpayment recoupment. CMS routinely conducts data analysis to identify potential problem areas and detect fraud, and audits important data fields used in our measures, including diagnosis and procedure codes and other elements that are consequential to payment.

Finally, we assess the reliability of the data elements by comparing model variable frequencies and odds ratios in 3 years of data.

Measure result reliability

The reliability of a measurement is the degree to which repeated measurements of the same entity agree with each other. For measures of hospital performance, the measured entity is naturally the hospital, and reliability is the extent to which repeated measurements of the same hospital give similar results. In line with this thinking, our approach to assessing reliability is to consider the extent to which assessments of a hospital using different but randomly selected subsets of patients produces similar measures of hospital performance. That is, we take a "test-retest" approach in which hospital performance is measured once using a random subset of patients, then measured again using a second random subset exclusive of the first, and the agreement of the two resulting performance measures compared across hospitals (Rousson et al., 2002).

For test-retest reliability, we combined index admissions from successive measurement periods into one dataset, randomly sampled half of patients within each hospital, calculated the measure for each hospital, and repeated the calculation using the second half. Thus, each hospital is measured twice, but each measurement is made using an entirely distinct set of patients. To the extent that the calculated measures of these two subsets agree, we have evidence that the measure is assessing an attribute of the hospital, not of the patients. As a metric of agreement we calculated the intra-class correlation coefficient (Shrout and Fleiss, 1979), and assessed the values according to conventional standards (Landis and Koch, 1977). Specifically, we used data Subsets A1 and A2, and calculated the RSRR for each hospital for each sample. The agreement of the two RSRRs was quantified for hospitals in each sample using the intra-class correlation as defined by ICC (2,1) by Shrout and Fleiss².

Using two independent samples provides an honest estimate of the measure's reliability, compared with using two random but potentially overlapping samples which would exaggerate the agreement. Moreover, because our final measure is derived using hierarchical logistic regression, and a known property of hierarchical logistic regression models is that smaller volume hospitals contribute less 'signal', a split sample using a single measurement period would introduce extra noise, potentially underestimating the actual test-retest reliability that would be achieved if the measure were reported using three years of data.

Rousson V, Gasser T, Seifert B. "Assessing intrarater, interrater and test-retest reliability of continuous measurements," *Statistics in Medicine*, 2002, 21:3431-3446.

Shrout P, Fleiss J. Intraclass correlations: uses in assessing rater reliability. *Psychological Bulletin*, 1979, 86, 420-3428.

Landis J, Koch G, The measurement of observer agreement for categorical data, *Biometrics*, 1977;33:159-174.

2a2.3 Testing Results (*Reliability statistics, assessment of adequacy in the context of norms for the test conducted*):

Data element reliability results

Overall, risk factor frequencies changed very little across the three-year period, and there were no notable differences in the odds ratios across years of data.

Measure result reliability results

There were 1,199,324 admissions in the combined three-year sample, with 599,723 in one randomly selected sample and 599,601 in the remaining sample. The agreement between the two RSMRs for each hospital was 0.406, which according to the conventional interpretation is "Moderate" (Landis & Koch, 1977). The intra-class correlation coefficient is based on a split sample of 3 years of data, resulting in a volume of patients in each sample equivalent to only 1.5 years of data, whereas the measure is likely to be reported with a full three years of data. Based on our experiences with similar measures using split sample, with 4 years (and volume equivalent to 2 years), the intra-class correlation coefficient would be even higher.

References:

1. Landis J, Koch G, The measurement of observer agreement for categorical data, *Biometrics*, 1977;33:159-174. The ICC for the 2007-2009 Subsets was 0.406.

2b. VALIDITY. Validity, Testing, including all Threats to Validity: H M L I

2b1.1 Describe how the measure specifications (measure focus, target population, and exclusions) are consistent with the evidence cited in support of the measure focus (criterion 1c) and identify any differences from the evidence:

N/A

2b2. Validity Testing. (Validity testing was conducted with appropriate method, scope, and adequate demonstration of validity.)

2b2.1 Data/Sample (Description of the data or sample including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities included):

We validated the administrative model with a medical-record based model when the measure was created. For the derivation of the medical record-based model, we used cases identified through a CMS quality initiative, which sampled admissions from fee-for-service Medicare beneficiaries for several clinical conditions, including pneumonia (Jencks et. al., 2000). Cases were identified over a 6-month period within each state, plus the District of Columbia and Puerto Rico, during the period July 1, 1998 through March 31, 1999. Based on the principal discharge diagnosis, approximately 750 pneumonia discharges per state were identified, and the corresponding medical records were abstracted by 2 clinical data abstraction centers (DynKePRO [York, PA] and FMAS Corporation [Rockville, MD]). In states with fewer than 750 pneumonia discharges, all cases were used. CMS subsequently conducted a re-measurement using the same data collection methodology for 2000 and 2001 discharges, and the combined 1998-2001 data that including a total of 75,616 cases served as the national pneumonia dataset for development of the medical record-based model (Jencks et. al., 2003).

References:

Jencks SF, Cuerdon T, Burwen DR, et al. Quality of medical care delivered to Medicare beneficiaries: A profile at state and national levels. *JAMA*. Oct 4 2000;284(13):1670-1676.

Jencks SF, Huff ED, Cuerdon T. Change in the quality of care delivered to Medicare beneficiaries, 1998-1999 to 2000-2001. *JAMA*. Jan 15 2003;289(3):305-312.

2b2.2 Analytic Method (Describe method of validity testing and rationale; if face validity, describe systematic assessment):

We sought to validate our best administrative pneumonia model (original model specification without the planned readmission algorithm) against the medical record model in the same cohort of patients for which pneumonia medical record data were available. We developed a measure cohort with the medical record data using the inclusion/exclusion criteria and risk-adjustment strategy that was consistent with the claims-based administrative measure but using chart-based risk adjusters, such as blood pressure, not available in the claims data. We then matched a sample of the same patients in the administrative data for comparison. We compared the output of the two measures, that is the state performance results, in the same group of patients.

ICD-9 to ICD-10 Conversion

Statement of Intent

- Convert measure to the new code set, but there are no changes to the measure.
- Take advantage of new specific code set for the measure with changes.
- The intent of the measure has changed.

Process of Conversion

We enlisted the help of clinicians with expertise in relevant areas to select and evaluate which ICD-10 codes map to the ICD-9 codes currently in use for this measure. The conversion of ICD-9 to ICD-10 is currently ongoing and the codes we have selected cannot yet be finalized since we lack sufficient ICD-10 data to evaluate the accuracy of coding/prevalence of ICD-10 codes. Once ICD-10 codes are officially in place and more data is available we will be able to provide a more accurate crosswalk.

2b2.3 Testing Results (Statistical results, assessment of adequacy in the context of norms for the test conducted; if face validity, describe results of systematic assessment):

See above

The performance of the administrative and medical record models is similar. The areas under the ROC curve are 0.63 and 0.59, respectively, for the two models. In addition, they are similar with respect to predictive ability. For the administrative model, the predicted readmission rate ranges from 8% in the lowest predicted decile to 30% in the highest predicted decile, a range of 22%. For the medical record model, the corresponding range is 10% to 26%, a range of 16%

We estimated hospital-level RSRRs using the corresponding hierarchical logistic regression administrative and medical record models for the linked patient sample. We then examined the linear relationship between the two sets of estimates using regression techniques and weighting by the total number of cases in each state. The correlation coefficient of the standardized rates from the administrative and medical record models is 0.96, and the proportion of the variance explained by the model is 0.92.

POTENTIAL THREATS TO VALIDITY. (All potential threats to validity were appropriately tested with adequate results.)

2b3. Measure Exclusions. (Exclusions were supported by the clinical evidence in 1c or appropriately tested with results demonstrating the need to specify them.)

2b3.1 Data/Sample for analysis of exclusions (Description of the data or sample including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities included):

We used all pneumonia admissions in 2007-2009 Medicare fee-for-service data (initial cohort which included 1,119,330 admissions) for the 65 and over model. We included 74,571 admissions in the 2006 all-payer California data for the 18 and over model.

2b3.2 Analytic Method (Describe type of analysis and rationale for examining exclusions, including exclusion related to patient preference):

All exclusions were determined by careful clinical review and have been used based on clinically relevant decisions. These exclusions are consistent with similar NQF-approved readmission measures.

2b3.3 Results (Provide statistical results for analysis of exclusions, e.g., frequency, variability, sensitivity analyses):

For the 65 and over model we examined overall frequencies and proportions of the admissions excluded for each exclusion criterion in all pneumonia admissions in 2007-2009 Medicare fee-for-service data (initial cohort included 1,119,330 admissions). The exclusion categories are not mutually exclusive.

1. In-hospital deaths (5.26%)
2. Transfer outs (0.50%)
3. Discharges against medical advice (AMA) (0.28%)
4. Hospitalizations without at least 30 days post-discharge information (0.60%)
5. Admissions within 30 days of a prior index admission (2.77%)

For the 18 and over model we examined overall frequencies and proportions of the admissions excluded for each exclusion criterion in all pneumonia admissions in 2006 all-payer California data (initial cohort included 74,571 admissions). The exclusion

categories are not mutually exclusive.

1. In-hospital deaths (5.20%)
2. Transfers to another acute care hospital (5.20%)
3. Discharges against medical advice (AMA) (1.20%)
4. Admissions within 30 days of a prior index admission (2.73%)

2b4. Risk Adjustment Strategy. (For outcome measures, adjustment for differences in case mix (severity) across measured entities was appropriately tested with adequate results.)

2b4.1 Data/Sample (Description of the data or sample including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities included):

Measure Development and Validation:

During initial measure development, using Medicare FFS beneficiaries age 65 and over, we tested the performance of the model (original model specification without the planned readmission algorithm) developed in a random selected half of the hospitalizations for pneumonia in data from the year 2006 (representing 226,545 cases discharged from 4,675 hospitals) against hospitalizations from the other half (representing 226,706 cases discharged from 4,671 hospitals).

Assessment of Temporal Trends in Model Performance Across Years of Data

For the 2007-2009 calendar year dataset, we reported results for each individual year as well as the 3-year combined results.

Application to Medicare FFS Beneficiaries Using Inpatient Data Only for Risk Adjustment

As part of testing the model in all-payer data, we also applied the model (original model specification without the planned readmission algorithm) to CMS data for Medicare FFS 65+ patients in California hospitals using only inpatient data for risk adjustment. Specifically, we created a 2006 measure cohort with complete one-year history data and 30-day follow-up data (N= 28,734).

Application to Patients Aged 18 and Older

We also applied the model (original model specification without the planned readmission algorithm) to all-payer data from California. The analytic sample included 74,571 cases aged 18 and older in the 2006 California Patient Discharge Data. When used in all-payer data, only admission claims data are used for risk adjustment, as the hospital discharge databases do not have outpatient claims.

2b4.2 Analytic Method (Describe methods and rationale for development and testing of risk model or risk stratification including selection of factors/variables):

Measure Development and Validation

This measure is fully risk-adjusted using a hierarchical logistic regression model to calculate hospital RSRRs accounting for differences in hospital case-mix. (See "risk adjustment methodology" for additional details.)

Approach to assessing model performance

During measure development, we computed five summary statistics for assessing model performance (Harrell and Shih 2001) for the development and validation cohort:

(1) over-fitting indices (over-fitting refers to the phenomenon in which a model accurately describes the relationship between predictive variables and outcome in the development dataset but fails to provide valid predictions in new patients)

(2) predictive ability

(3) area under the receiver operating characteristic (ROC) curve

(4) distribution of residuals

(5) model chi-square (A test of statistical significance usually employed for categorical data to determine whether there is a good fit between the observed data and expected values; i.e., whether the differences between observed and expected values are attributable to true differences in characteristics or instead the result of chance variation).

We tested the performance of the model developed in a random selected half of the hospitalizations for pneumonia in data from the year 2006 against hospitalizations from the other half.

Assessment of Temporal Trends in Model Performance Across Years of Data

Across years, we examined consistency in frequency of risk-adjustment variables and parameter estimates for risk-adjustment variables and model performance (C statistic).

Application to Medicare FFS Beneficiaries Using Inpatient Data Only for Risk Adjustment

To help determine whether the measure could be applied to Medicare FFS 65+ patients using only Medicare Part A data, we performed analyses to assess how the model performs when using only inpatient claims data for risk adjustment, as all-payer hospital discharge databases do not have outpatient claims. To assess the validity of using only admission claims data for risk adjustment, we fit the model separately using the full data and using only admission claims data and (a) compared the odds ratios (ORs) for the various risk factors; (b) conducted a reclassification analysis to compare risk prediction at the patient level; (c) compared model performance in terms of the c-statistic (discrimination); and (d) compared hospital-level risk-standardized rates (scatterplot, correlation coefficient, and R2) to assess whether the model with only admission claims data is different from the current model in profiling hospital rates.

Application to Patients Aged 18 and Older

To help determine whether the measure could be applied to a population of patients aged 18+, we examined the interaction terms between age (18-64 vs. 65+) and each of the other risk factors. Specifically, we fit the model in all patients 18+ with and without interaction terms and (a) conducted a reclassification analysis to compare risk prediction at the patient level; (b) compared the c-statistic; and (c) compared hospital-level risk-standardized rates (scatterplot, correlation coefficient, and R-square) to assess whether the model with interactions is different from the current model in profiling hospital rates.

Reference:

Harrell FE, Shih YCT. Using full probability models to compute probabilities of actual interest to decision makers. *Int J Technol Assess Health Care*. 2001;17:17-26.

2b4.3 Testing Results (*Statistical risk model: Provide quantitative assessment of relative contribution of model risk factors; risk model performance metrics including cross-validation discrimination and calibration statistics, calibration curve and risk decile plot, and assessment of adequacy in the context of norms for risk models. Risk stratification: Provide quantitative assessment of relationship of risk factors to the outcome and differences in outcomes among the strata*):

Measure Development and Validation

The performance was not substantively different in the validation sample (ROC area = 0.63) compared with the development sample (2006). The models appear well calibrated, with over-fitting indices of (-0.002, 0.997).

Residuals lack of fit (<2, [-2,0],[0,2],[2+]): 0.00, 82.55, 7.45, 9.99 Model Chi-square [# of covariates]: 6,870 [40]

Predictive ability (lowest decile %, highest decile %): 9.0%, 31%

Area under ROC curve: 0.63

For the validation cohort the results are summarized below:

Residuals lack of fit (<2, [-2,0],[0,2],[2+]): 0.00, 82.67, 7.31, 10.03 Model Chi-square [# of covariates]: 16,241 [40]

Predictive ability (lowest decile %, highest decile %): 8.0%, 31%

Area under ROC curve: 0.63

Assessment of Temporal Trends in Model Performance Across Years of Data

The frequency of risk-adjustment variables and parameter estimates for risk-adjustment variables and model performance was stable over all time periods.

Model Performance in Medicare FFS Beneficiaries Using Inpatient Data Only for Risk

Adjustment using CMS data for Medicare FFS 65+ beneficiaries in California hospitals: (a) the magnitude of odds ratios for most risk factors was similar when comparing the model using full data and using only admission claims data; (b) when comparing the model with full data and with only admission claims data, the reclassification analysis demonstrated good patient-level risk prediction; (c) the c-statistic was similar (0.632vs. 0.628); and (d) hospital-level risk-standardized rates were highly correlated (ICC=0.985).

Model Performance in Patients Aged 18 and Older

When the model was applied to all patients 18 and over (18+), overall discrimination was good (c-statistic=0.666). In addition, there was good discrimination and predictive ability in both those aged 18-64 and those aged 65+. Moreover, the distribution of Pearson residuals was comparable across the patient subgroups. When comparing the model with and without interaction terms, (a) the reclassification analysis demonstrated that nearly all patients were found to be in a similar risk category; (b) the c-statistic was nearly identical (0.669vs. 0.666); and (c) hospital-level risk-standardized rates were highly correlated (ICC=0.997). Thus, the inclusion of the interactions did not substantively affect either patient-level model performance or hospital-level results.

Therefore, the measure can be applied to all payer data for patients 18 and older.

References:

Krumholz HM, Normand S-LT, Galusha DH, Mattera JA, Rich AS, Wang YF, Wang Y. et al. Risk-Adjustment Models for AMI and HF: 30-Day Mortality: Report prepared for the Centers for Medicare & Medicaid Services; 2005. Available at: <http://www.qualitynet.org/>

Bernheim SM, et al. 2010 Measures Maintenance Technical Report: Acute Myocardial Infarction, Heart Failure and Pneumonia 30-day Risk Standardized Mortality Rate. 2010 Available at: <http://www.qualitynet.org/dcs/ContentServer?c=Page&pagename=QnetPublic/Page/QnetTier3&cid=1163010421830>

2b4.4 If outcome or resource use measure is not risk adjusted, provide rationale and analyses to justify lack of adjustment: N/A

2b5. Identification of Meaningful Differences in Performance. *(The performance measure scores were appropriately analyzed and discriminated meaningful differences in quality.)*

2b5.1 Data/Sample *(Describe the data or sample including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities included):*

The data are based on RSRRs calculated for pneumonia hospitalization among Medicare FFS patients aged 65+ from July 2008 to June 2011 (applying the new planned readmission algorithm), and include 1,096,708 hospitalizations from 4,859 hospitals.

2b5.2 Analytic Method *(Describe methods and rationale to identify statistically significant and practically/meaningfully differences in performance):*

Below we present the distribution of the current measure.

For public reporting of the measure, CMS characterizes the uncertainty associated with the RSRR by estimating the 95% interval estimate. This is similar to a 95% confidence interval but is calculated differently. If the RSRR's interval estimate does not include the national crude readmission rate (is lower or higher than the rate), then CMS is confident that the hospital's RSRR is different from the national rate, and describes the hospital on the Hospital Compare website as "better than the U.S. national rate" or "worse than the U.S. national rate." If the interval includes the national rate, then CMS describes the hospital's RSRR as "no different than the U.S. national rate" or "the difference is uncertain." CMS does not classify performance for hospitals that have fewer than 25 pneumonia cases in the three-year period.

2b5.3 Results (Provide measure performance results/scores, e.g., distribution by quartile, mean, median, SD, etc.; identification of statistically significant and meaningful differences in performance):

There is substantial variation in pneumonia RSRRs among hospitals. The median hospital RSRR for pneumonia was 17.7% with a range of 13.4% to 25.5%. The 5th percentile was 15.7% and the 95th percentile was 20.5%. The interquartile range was 16.9% to 18.7%.

2b6. Comparability of Multiple Data Sources/Methods. (If specified for more than one data source, the various approaches result in comparable scores.)

2b6.1 Data/Sample (Describe the data or sample including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities included):

The measure performs well in both Medicare FFS data and all-payer data.

2b6.2 Analytic Method (Describe methods and rationale for testing comparability of scores produced by the different data sources specified in the measure):

See attached all-payer report

2b6.3 Testing Results (Provide statistical results, e.g., correlation statistics, comparison of rankings; assessment of adequacy in the context of norms for the test conducted):

See attached all-payer report

2c. Disparities in Care: H M L I NA (If applicable, the measure specifications allow identification of disparities.)

2c.1 If measure is stratified for disparities, provide stratified results (Scores by stratified categories/cohorts): Measure is not stratified for disparities.

2c.2 If disparities have been reported/identified (e.g., in 1b), but measure is not specified to detect disparities, please explain:

The analyses performed by CMS (described in section 1b) demonstrate that hospitals have similar and overlapping performance on the measure regardless of the proportion of patients of low socioeconomic status or of African-American race. Importantly, the analyses show that hospitals with high proportions of low socioeconomic status patients or high proportions of African-American patients are able to perform well on the measure. For this reason CMS does not plan to stratify the measure.

2.1-2.3 Supplemental Testing Methodology Information:

Attachment

All-Payer Testing Report_AMI_HF_PN_Measures_Final [1.17.12]-634623951503039268.pdf

Steering Committee: Overall, was the criterion, Scientific Acceptability of Measure Properties, met?
(Reliability and Validity must be rated moderate or high) Yes No

Provide rationale based on specific subcriteria:

If the Committee votes No, STOP

3. USABILITY

Extent to which intended audiences (e.g., consumers, purchasers, providers, policy makers) can understand the results of the measure and are likely to find them useful for decision making. (**evaluation criteria**)

C.1 Intended Actual/Planned Use (Check all the planned uses for which the measure is intended): [Public Reporting, Quality Improvement with Benchmarking](#) (external benchmarking to multiple organizations)

3.1 Current Use (Check all that apply; for any that are checked, provide the specific program information in the following questions): [Public Reporting, Quality Improvement](#) (Internal to the specific organization)

3a. Usefulness for Public Reporting: H M L I

(The measure is meaningful, understandable and useful for public reporting.)

3a.1. Use in Public Reporting - disclosure of performance results to the public at large (If used in a public reporting program, provide name of program(s), locations, Web page URL(s)). If not publicly reported in a national or community program, state the reason AND plans to achieve public reporting, potential reporting programs or commitments, and timeline, e.g., within 3 years of endorsement: [**For Maintenance** – If not publicly reported, describe progress made toward achieving disclosure of performance results to the public at large and expected date for public reporting; provide rationale why continued endorsement should be considered.]

The measure has been publicly reported on Hospital Compare (www.hospitalcompare.hhs.gov) since June 2009 and is used in CMS' Hospital Inpatient Quality Reporting Program (Formerly RHQDAPU).

3a.2. Provide a rationale for why the measure performance results are meaningful, understandable, and useful for public reporting. If usefulness was demonstrated (e.g., focus group, cognitive testing), describe the data, method, and results: The measure has been systematically evaluated by a group of clinical experts throughout the development process to construct a measure for use in public reporting. We have received input and feedback on key issues related to the meaningfulness, usefulness, and design of the measure. Meetings were held throughout the development process and we received input and feedback on key methodological and clinical decisions to ensure the measure is meaningful and useful.

In addition, similar measures for acute myocardial infarction (AMI) and heart failure underwent consumer testing prior to being publicly reported and were found to be useful for publicly reporting outcomes.

3.2 Use for other Accountability Functions (payment, certification, accreditation). If used in a public accountability program, provide name of program(s), locations, Web page URL(s): The measure has been publicly reported on Hospital Compare (www.hospitalcompare.hhs.gov) since June 2009 and is used in CMS' Hospital Inpatient Quality Reporting Program (Formerly RHQDAPU).

3b. Usefulness for Quality Improvement: H M L I

(The measure is meaningful, understandable and useful for quality improvement.)

3b.1. Use in QI. If used in quality improvement program, provide name of program(s), locations, Web page URL(s): [**For Maintenance** – If not used for QI, indicate the reasons and describe progress toward using performance results for improvement].

The measure has been publicly reported on Hospital Compare (www.hospitalcompare.hhs.gov) since June 2009 and is used in CMS' Hospital Inpatient Quality Reporting Program (Formerly RHQDAPU).

3b.2. Provide rationale for why the measure performance results are meaningful, understandable, and useful for quality improvement. If usefulness was demonstrated (e.g., QI initiative), describe the data, method and results:

A hospital-level, 30-day readmission measure for pneumonia patients may incentivize hospitals to improve quality of care for this high-risk population.

Overall, to what extent was the criterion, Usability, met? H M L I

Provide rationale based on specific subcriteria:

4. FEASIBILITY

Extent to which the required data are readily available, retrievable without undue burden, and can be implemented for performance measurement. (**evaluation criteria**)

4a. Data Generated as a Byproduct of Care Processes: H M L I

4a.1-2 How are the data elements needed to compute measure scores generated? (Check all that apply).

Data used in the measure are:

Coded by someone other than person obtaining original information (e.g., DRG, ICD-9 codes on claims)

4b. Electronic Sources: H M L I

4b.1 Are the data elements needed for the measure as specified available electronically (Elements that are needed to compute measure scores are in defined, computer-readable fields): ALL data elements in electronic claims

4b.2 If ALL data elements are not from electronic sources, specify a credible, near-term path to electronic capture, OR provide a rationale for using other than electronic sources:

4c. Susceptibility to Inaccuracies, Errors, or Unintended Consequences: H M L I

4c.1 Identify susceptibility to inaccuracies, errors, or unintended consequences of the measurement identified during testing and/or operational use and strategies to prevent, minimize, or detect. If audited, provide results:

This measure uses variables from claims data submitted by hospitals for payment. Prior research has demonstrated that administrative claims data can be used to develop risk-adjusted outcomes measures for both mortality and readmission following hospitalization for acute myocardial infarction (Krumholz et al., 2006a; Krumholz et al., 2011), heart failure (Krumholz et al., 2006b; Keenan et al., 2008), and pneumonia (Bratzler et al., 2011; Lindenauer et al., 2011), and that the models produce estimates of risk-standardized rates that are very similar to rates estimated by models based on medical record data. This high level of agreement supports the use of the claims-based risk-adjusted models for public reporting. The models have also demonstrated consistent performance across years of claims data.

The approach to gathering risk factors for patients also mitigates the potential limitations of claims data. Because not every diagnosis is coded at every visit, we use inpatient, outpatient, and physician claims data for the year prior to admission, and diagnosis codes during the index admission, for risk adjustment when the measure is used in Medicare FFS data. When the measure is used in all-payer data, only admission claims data (from the index hospitalization and prior year) are used for risk adjustment; however, model testing demonstrated both strong patient-level model performance and consistent hospital-level results when using only admission claims data. The 1-year time frame provides a more comprehensive view of patients' medical histories than is provided by the secondary diagnosis codes from the index hospitalization alone. If a diagnosis appears in some visits and not others, it is included, minimizing the effect of incomplete coding. We were careful, however, to include information about each patient's status at admission and not to adjust for possible complications of the admission. Although some codes, by definition, represent conditions that are present before admission (e.g. cancer), other codes and conditions cannot be differentiated from complications during the hospitalization (e.g. infection or shock). If these are secondary diagnoses from the index admission, then they are not adjusted for in the analysis.

References:

Krumholz HM, Wang Y, Mattera JA, Wang Y-F, Han LF, Ingber MJ, Roman S, Normand SL (2006a). An administrative claims model suitable for profiling hospital performance based on 30-day mortality rates among patients with an acute myocardial infarction. *Circulation*. 2006 Apr 4;113(13):1683-92.

Krumholz HM, Lin Z, Drye EE, Desai MM, Han LF, Rapp MT, Mattera JA, Normand SL. An administrative claims measure suitable for profiling hospital performance based on 30-day all-cause readmission rates among patients with acute myocardial infarction. *Circulation: Cardiovascular Quality and Outcomes*. 2011 Mar 1;4(2):243-52.

Krumholz HM, Wang Y, Mattera JA, Wang Y-F, Han LF, Ingber MJ, Roman S, Normand SL (2006b). An administrative claims model suitable for profiling hospital performance based on 30-day mortality rates among patients with heart failure. *Circulation*. 2006 Apr 4;113(13):1693-701.

Keenan PS, Normand SL, Lin Z, Drye EE, Bhat KR, Ross JS, Schuur JD, Stauffer BD, Bernheim SM, Epstein AJ, Wang Y-F, Herrin J, Chen J, Federer JJ, Mattera JA, Wang Y, Krumholz HM. An administrative claims measure suitable for profiling hospital performance on the basis of 30-day all-cause readmission rates among patients with heart failure. *Circulation: Cardiovascular Quality and Outcomes*. 2008 Sep;1(1):29-37.

Bratzler DW, Normand SL, Wang Y, O'Donnell WJ, Metersky M, Han LF, Rapp MT, Krumholz HM. An administrative claims model for profiling hospital 30-day mortality rates for pneumonia patients. *Public Library of Science One*. 2011 Apr 12;6(4):e17401.

Lindenauer PK, Normand SL, Drye EE, Lin Z, Goodrich K, Desai MM, Bratzler DW, O'Donnell WJ, Metersky ML, Krumholz HM. Development, validation, and results of a measure of 30-day readmission following hospitalization for pneumonia. *Journal of Hospital Medicine*. 2011 Mar;6(3):142-50.

4d. Data Collection Strategy/Implementation: H M L I

A.2 Please check if either of the following apply (regarding proprietary measures):

4d.1 Describe what you have learned/modified as a result of testing and/or operational use of the measure regarding data collection, availability of data, missing data, timing and frequency of data collection, sampling, patient confidentiality, time and cost of data collection, other feasibility/implementation issues (e.g., fees for use of proprietary measures):
 Administrative data is routinely collected as part of the billing process.

Overall, to what extent was the criterion, *Feasibility*, met? H M L I

Provide rationale based on specific subcriteria:

OVERALL SUITABILITY FOR ENDORSEMENT

Does the measure meet all the NQF criteria for endorsement? Yes No

Rationale:

If the Committee votes No, STOP.

If the Committee votes Yes, the final recommendation is contingent on comparison to related and competing measures.

5. COMPARISON TO RELATED AND COMPETING MEASURES

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

5.1 If there are related measures (either same measure focus or target population) or competing measures (both the same measure focus and same target population), list the NQF # and title of all related and/or competing measures:

0330 : Hospital 30-day, all-cause, risk-standardized readmission rate (RSRR) following heart failure hospitalization

0468 : Hospital 30-day, all-cause, risk-standardized mortality rate (RSMR) following pneumonia hospitalization

0505 : Hospital 30-day all-cause risk-standardized readmission rate (RSRR) following acute myocardial infarction (AMI) hospitalization.

0708 : Proportion of Patients Hospitalized with Pneumonia that have a Potentially Avoidable Complication (during the Index Stay or in the 30-day Post-Discharge Period)

5a. Harmonization

5a.1 If this measure has EITHER the same measure focus OR the same target population as NQF-endorsed measure(s):
 Are the measure specifications completely harmonized? Yes

5a.2 If the measure specifications are not completely harmonized, identify the differences, rationale, and impact on interpretability and data collection burden:

5b. Competing Measure(s)

5b.1 If this measure has both the same measure focus and the same target population as NQF-endorsed measure(s): Describe why this measure is superior to competing measures (e.g., a more valid or efficient way to measure quality); OR provide a rationale for the additive value of endorsing an additional measure. (Provide analyses when possible):
N/A

CONTACT INFORMATION

Co.1 Measure Steward (Intellectual Property Owner): Centers for Medicare and Medicaid Services, 7500 Security Boulevard , Mail Stop S3-01-02, Baltimore, Maryland, 21244-1850

Co.2 Point of Contact: Lein, Han, Ph.D., Government Task Leader, Lein.han@cms.hhs.gov, 410-786-0205-

Co.3 Measure Developer if different from Measure Steward: Centers for Medicare & Medicaid Services, 7500 Security Boulevard, Baltimore, Maryland, 21244

Co.4 Point of Contact: Lein, Han, Lein.han@cms.hhs.gov, 410-786-6738-

Co.5 Submitter: Kanchana, Bhat, kanchana.bhat@yale.edu, 410-786-6738-, Centers for Medicare & Medicaid Services

Co.6 Additional organizations that sponsored/participated in measure development:
MPR: Mathematica Policy Research; RTI-Research Triangle Institute

Co.7 Public Contact: Lein, Han, Ph.D., Government Task Leader, Lein.han@cms.hhs.gov, 410-786-0205-, Centers for Medicare & Medicaid Services

ADDITIONAL INFORMATION

Workgroup/Expert Panel involved in measure development

Ad.1 Provide a list of sponsoring organizations and workgroup/panel members' names and organizations. Describe the members' role in measure development.

The working group involved in the initial measure development is detailed in the original technical report available at www.qualitynet.org

Ad.2 If adapted, provide title of original measure, NQF # if endorsed, and measure steward. Briefly describe the reasons for adapting the original measure and any work with the original measure steward:

Measure Developer/Steward Updates and Ongoing Maintenance

Ad.3 Year the measure was first released: 2009

Ad.4 Month and Year of most recent revision: 08, 2011

Ad.5 What is your frequency for review/update of this measure? Yearly

Ad.6 When is the next scheduled review/update for this measure? 07, 2012

Ad.7 Copyright statement: N/A

Ad.8 Disclaimers:

Ad.9 Additional Information/Comments: www.qualitynet.org for Measure Methodology report and Maintenance reports

Date of Submission (MM/DD/YY): 10/18/2011

NQF #0506 Hospital 30-day, all-cause, risk-standardized readmission rate (RSRR) following pneumonia hospitalization, Last Updated Date: Oct 15, 2012