



TO: National Quality Forum and Pulmonary Steering Committee

FROM: Bob Rehm, Assistant Vice President, Performance Measurement
Ben Hamlin, Director, Performance Measurement
Dawn Alayon, Senior Health Care Analyst, Performance Measurement

DATE: May 30, 2012

RE: Request for Reconsideration of Measure 0549: Pharmacotherapy Management of COPD Exacerbation (PCE)

On March 21, 2012, the National Quality Forum (NQF) Pulmonary Steering Committee (SC) reviewed the currently endorsed NQF measure 0549: Pharmacotherapy Management of COPD Exacerbation (PCE), an administrative claims-based health plan measure assessing the percentage of COPD exacerbations for members 40 years of age and older for which appropriate medications were dispensed.

Two rates are reported.

1. Dispensed a systemic corticosteroid within 14 days of the event
2. Dispensed a bronchodilator within 30 days of the event

Measure 0549 was the 22nd of 23 measures evaluated during the day and following strong affirmative votes in the three sub-criteria of Importance and the first Scientific Acceptability sub-criterion of Reliability, the Pulmonary Steering Committee's next vote on the sub-criterion of Validity received a very close vote resulting in suspension of further discussion and moving on to the last measure of the day. Given the volume of measures reviewed, the nature of the questions raised by the SC, and the observed and self-characterized fatigue on the part of the SC, we believe additional consideration at this point of the SC review, with leadership and guidance from the NQF staff would have been appropriate for a measure in broad national use and publicly reported. We have observed in several other steering committees that NQF leadership has offered additional guidance when measures have unexpectedly hit a snag, or a vote is close.

SC Voting Record for NQF Measure 0549 March 21, 2012

Importance	H	M	L	I	
1a – High Impact	15	3	0	0	Pass 18-0
1b – Performance Gap	2	13	2	1	Pass 15-3
	Y	N	I		
1c – Evidence to support measure	15	1	2		Pass 16-2
Scientific Acceptability	H	M	L	I	
2a - Reliability	1	11	5	1	Pass 12-6
2b - Validity	0	7	8	2	Not Pass- review terminated
Usability	H	M	L	I	
3a-b					No vote
Feasibility	H	M	L	I	
4a-d					No vote
Overall Suitability	Y	N			
Pass/Fail					No vote

The following morning, NCQA provided additional information in response to questions that were raised during the SC review and requested that NQF staff review both the vote and the discussion during the measure review and reconsider the close SC vote on Validity. In response NQF leadership directly informed the SC that they would be receiving additional information from NCQA that may clarify SC concerns. Thirty days later

following an additional request for follow-up, NQF staff responded that a large majority of SC members reaffirmed the original evaluation on the measure. The Committee appreciated the additional information provided; however, they did not feel that the information addressed the issues they had raised.

In the intervening period since the SC meeting, NCQA reviewed the technical draft report, the WG and SC audio recordings and transcripts. Based on this review, we found that SC members were discussing issues unrelated to the sub-criteria under review. NCQA would like this opportunity to address those issues and respectfully requests that the NQF SC reconsider the measure based on the clarifications provided.

Several questions about the measure logic were posed by the SC during their deliberations on the Importance sub-criteria including:

1. ***Does 0549 capture samples providing in the ED or hospital?*** There are currently no mechanisms for capturing the provision of samples in any setting, whether that is a health plan, a hospital, ED or physician office. NCQA knows of no NQF-endorsed measure that includes a specification that includes capturing of drug samples. In contrast to many NQF-endorsed measures that use “prescribed” data, all NCQA medication-related measures rely solely on dispensed drug information. A survey of several prominent ED physicians was conducted following this meeting and there was general agreement that this practice is almost non-existent for the types of medications required for this measure.
2. ***Does 0549 capture prescriptions dispensed in the ED?*** Yes-ED medication events are submitted by the hospital/ED and paid by the health plan and therefore captured in the paid pharmacy claims data, the same way as any dispensed ambulatory medications.
3. ***How does 0549 capture listed medications that are in current use (active prescription) at the time of the event? Does 0549 require a look back period?*** Health plans look back for active prescriptions. If a health plan member has a valid prescription that persists through the date of the exacerbation this is used to count the person as numerator compliant. Health plans reporting this measure are positively rewarded for capturing active prescriptions that indicate numerator compliance. With over 75 HEDIS measures reported by health plans, NCQA has evolved measure specifications that are feasible by a wide range of reporting entities, are audited and have demonstrated broad acceptance in the marketplace.

During the discussion period for Importance and all its sub-criteria, the SC discussion focused exclusively on the sub-criteria of Validity with no further discussion of this measure’s high impact, performance gap, and evidence. Prior to the Validity vote, Dr. Stephen Grossbart, the SC Co-Chair, commented that this sub-criterion received a moderate vote from the WG and also noted that the Validity concerns raised during the prior discussion were not raised during the initial review. As reflected in the NQF measure submission form, the field testing demonstrated that this measure was able to capture COPD exacerbations in administrative claims data, validating the denominator data elements (i.e. patients with COPD exacerbations) and the numerator data elements (i.e. patients treated with inhaled bronchodilator while in the ED and patients given systemic corticosteroids while in the ED).

Our review of the SC discussions captured by meeting recordings and the subsequent voting concludes that the SC members discussion prior to each sub-criterion section and vote were focused on issues not related to the specific sub-criterion and therefore we are concerned that the SC may not have been evaluating measure 0549 against the criteria for which they were voting. This deviation from normal steering committee practice resulted in this measure not continuing for the SC vote on Usability, Feasibility and overall suitability for continued NQF endorsement. Combined with the other issues raised, NCQA believes that due to these process issues, measure 0549 warrants further review and consideration.

Addendum: NCQA Measure, Data Submission and Auditing Processes

Background: NQF measure #0549 has been part of the HEDIS measure set since 2008. Because of NCQA's stringent data collection and auditing process, the data submitted to NCQA for HEDIS Health Plan reporting is of the highest quality and validity. NCQA has evaluated the measure for face validity through the Respiratory Measurement Advisory Panel (RMAP), an advisory body of clinicians and methodological experts (Dr. Kevin Weiss served as Chair of this panel from 2006-2009 and was involved in the initial development and testing of measure 0549). The RMAP includes representatives from key stakeholder groups, including the CDC, pulmonologists, pharmacologists, health service researchers and healthcare delivery organizations. Following NCQA's 30 day public comment period, review of the field testing results and RMAP recommendations, the Committee on Performance Measurement (CPM) approved the specification for inclusion in HEDIS.

Once a measure is included in the HEDIS Health Plan measurement set, NCQA requires a rigid data submission and auditing process for all HEDIS measures.

HEDIS Data Submission Process: NCQA collects HEDIS data directly from Health Plan Organizations and Preferred Provider Organizations for multiple purposes via the Health Organization Questionnaire (HOQ) and HEDIS non-survey data through the Interactive Data Submission System (IDSS).

As an administrative claims-based, health plan measure, this measure can only capture medications that have been dispensed.

NCQA Auditing Process: NCQA recognizes that, despite the clear specifications defined for HEDIS measures, data collection and calculation methods may vary, and other errors may taint the results, diminishing the usefulness of HEDIS data for managed care organization (MCO) comparison. In order for HEDIS to reach its full potential, NCQA conducts an independent audit of HEDIS collection and reporting processes, as well as an audit of the data which are manipulated by those processes, in order to verify that HEDIS specifications are met. NCQA has developed a precise, standardized methodology for verifying the integrity of HEDIS collection and calculation processes through a two-part program consisting of an overall information systems capabilities assessment (IS standards) followed by an evaluation of the MCO's ability to comply with HEDIS specifications (HD standards). NCQA-certified auditors using standard audit methodologies help enable purchasers to make more reliable "apples-to-apples" comparisons between health plans.

The HEDIS Compliance Audit addresses the following functions:

- 1) information practices and control procedures
- 2) sampling methods and procedures
- 3) data integrity
- 4) compliance with HEDIS specifications
- 5) analytic file production
- 6) reporting and documentation

June 5, 2012

Interim President and CEO
National Quality Forum
600 13th Street, NW, Suite 500 North
Washington, DC 20005

SUBJECT: Comments Regarding the Pulmonary Maintenance Review Project

Dear Interim CEO:

On behalf of our more than 5,000 member hospitals, health systems and other health care organizations, and our 42,000 individual members, the American Hospital Association (AHA) appreciates the opportunity to comment on the National Quality Forum's (NQF) Pulmonary Maintenance Review project. Reviewing measures that have previously been endorsed is one of the fundamental tenants of measure development and we commend the NQF for continually holding all developers to this high standard. Recognizing the current state of science and medicine and how advances alter our fundamental delivery of care is something that measures must be responsive to. Our comments on the measures under review are included below.

Hospital 30-day, all cause, risk-standardized measures

There are a total of four measures in this pulmonary maintenance review project that are developed by Yale:

1. Maintenance Review: Thirty-day all-cause risk standardized readmission rate following pneumonia hospitalizations (0506);
2. Maintenance Review: Thirty-day all-cause risk standardized mortality rate following pneumonia hospitalizations (0468);
3. First-time Endorsement: Thirty-day all-cause risk standardized readmission rate following COPD hospitalizations (1891); and
4. First-time Endorsement: Thirty-day all-cause risk standardized mortality rate following COPD hospitalizations (1893).

Failure to adjust for factors beyond the hospital's control. These four measures, along with all of Yale's 30-day all cause risk standardized measures, consistently fail to adjust for factors beyond a hospital's control. Changes need to be made to these measures to properly stratify for patient characteristics (dual eligible status and race/ethnicity) and exclude extreme circumstances (transplant, end-stage renal disease, burn, trauma, psychosis and substance abuse). Further, these measures must address patient compliance with treatment plans and the quality of post-acute care provided to the patient.

The fact that these measures are "all-cause" means they will include readmissions and deaths that have nothing to do with the specific condition associated with the measure. A COPD patient could be readmitted/die from a fall, a car accident, or an unrelated condition not risk-adjusted in the measure. However, that readmission/death will still be included against the hospital in these measures. For these reasons, the hospital 30-day, all cause, risk-standardized measures

that are publicly reported provide information that is unsuitable for medical decision-making by consumers, because patients cannot parse out the many variables necessary to truly assess a hospital's performance.

Reliability. On February 13 CMS released a reliability study^A, required by the ACA, for claims-based measures. The study shows that the majority of claims-based measures currently used in CMS's programs are unreliable. Among the measures included in the study was the pneumonia mortality measure. The CMS study states that "reliability of an outcome measure is the extent to which variation in the measure is due to variation in quality of care rather than random variation due to the sample of cases observed. The statistical concept of reliability (R) used to determine the minimum case size for a particular measure is whether a hospital's ranking on that measure, compared to its performance in other periods or compared to other hospitals, is likely to be the same if we take repeated samples of the hospital's own cases. R depends on the rate's variance between hospitals, the variance of the rate within a hospital's own cases, and the number of discharges from a given hospital." Table 1 below includes the information from the study on the pneumonia mortality measure.

Table 1 – Reliability of Pneumonia Mortality Measure

Measure	6 Months		12 Months		18 Months		24 Months	
	Median* Reliability	% of Hospitals $R \geq 0.4^{**}$	Median* Reliability	% of Hospitals $R \geq 0.4^{**}$	Median* Reliability	% of Hospitals $R \geq 0.4^{**}$	Median* Reliability	% of Hospitals $R \geq 0.4^{**}$
PN Mortality (R = 0.4 with 211 cases)	0.11	1	0.19	8	0.27	22	0.32	35
* Reliability of measure of hospital of median case size								
**Proportion of hospitals with case size large enough that $R \geq 0.4$								

The CMS study indicates that "R = 0.4 is considered to be the lower limit of 'moderate' reliability." Yale sets its reliability rates at 0.4 for all of its hospital 30-day, all cause, risk-standardized measures. "The lower limit of moderate reliability" is not sufficient for NQF-endorsement. CMS's standard^B for non-claims-based measures is R = 0.75. Further, we note that in order to achieve R = 0.4 for the pneumonia mortality measure, CMS recommends that a minimum of 211 cases are needed. The current minimum threshold for the Yale pneumonia measure is 25 cases—a major difference. Though the measure developer uses 36-months of data and this study does not address 36-months of data, we still do not believe this measure is reliable. **At a minimum, we urge the steering committee to request this reliability data of the measure developer for both the pneumonia and COPD mortality measures.**

KNG, an independent research firm, has replicated^B CMS's reliability study for both the pneumonia and COPD readmission measures. The results of this analysis are included in Table 2, below.

Table 2 – Reliability of Pneumonia and COPD Readmission Measures

Measure	12 Months		24 Months		36 Months	
	Median* Reliability	% of Hospitals R ≥ 0.4**	Median* Reliability	% of Hospitals R ≥ 0.4**	Median* Reliability	% of Hospitals R ≥ 0.4**
PN Readmission (R = 0.4 with 121 cases)	0.28	21%	0.43	57%	0.53	75%
PN Readmission (R = 0.7 with 421 cases)		0%		5%		16%
COPD Readmission (R = 0.4 with 154 cases)	0.22	14%	0.37	44%	0.46	62%
COPD Readmission (R = 0.7 with 537 cases)		0%		2%		9%
* Reliability of measure of hospital of median case size						
**Proportion of hospitals with case size large enough that R ≥ 0.4						

This data illustrates that both readmission measures fail to reach the industry standard of $R = 0.75$. KING recommends that a minimum of 121 cases are needed for the pneumonia measure and 154 cases are needed for the COPD measure in order to achieve $R = 0.4$. The current minimum threshold for the Yale readmissions measure is 25 cases—a major difference.

Pneumonia and COPD readmission measures

The *Affordable Care Act* calls for these readmission measures to recognize planned cases and unrelated admissions; however these measures do not. Therefore, they may be erroneously counting certain admissions and readmissions. Because this readmission measure is currently publicly reported and will be used to change reimbursement for hospitals in the near future, this omission must be rectified immediately. We urge the steering committee to take this into consideration and we ask the steering committee to query the measure developer on why this change has not been made to the measure.

Harmonization. Though the pneumonia and COPD readmission measures do not properly account for unrelated readmissions, we note that another NQF-endorsed measure (30-day all-cause all-condition risk standardized measure) does remove some planned readmissions. Therefore, pneumonia and COPD measures are not harmonized with the all-condition readmission measure. The all-condition readmission measure excludes patients undergoing medical treatment for their cancer as well as 27 other categories of procedures that are considered unplanned. **We urge the steering committee to require the measure developer to harmonize these methodologies by also excluding these planned readmissions from the pneumonia and COPD readmission measures.**

COPD mortality measure

Based on comments in the “Evidence” section of the steering committee’s evaluation for this measure, it appears that the goal of this measure is to encourage hospitals to follow guidelines for COPD exacerbation management. However, the measure developer could create a more direct, process of care measure that evaluates whether the hospital followed such guidelines, or that measures their coordination of care, rather than creating a measure impacted by multiple outside factors.

The exclusion of FFS Medicare patients enrolled in Hospice is a good attempt to delineate between patients going through aggressive treatment and those who have chosen to enter palliative care programs. It is not clear whether there are Medicare patients who are not part of this exclusion. In other words, why does the measure exclude FFS Medicare patients enrolled in Hospice as opposed to all Medicare patients enrolled in Hospice? In addition, this exclusion does not completely remove from the measure population all of those patients who are terminally ill, because Hospice programs may be full. It also does not appear to remove sicker patients who have signed a DNR measure. In short, the risk adjustment and exclusions do not seem to identify and address the sickest patients.

The exclusions for this measure lack consistency with the exclusions for the 30-day all-cause risk-standardized mortality rate for pneumonia. The PN measure includes an exclusion for patients discharged alive on the day of admission or the following day who did not get transferred, because it is unlikely the patient had a significant pneumonia diagnosis. Why would the same kind of exclusion, such as lack of a significant COPD episode, not be part of the 30-day COPD measure?

The description of the adjustment/stratification for this measure included in the draft report also notes that “the RSMR is calculated as the ratio of the number of “predicted” to the number of “expected” deaths, multiplied by the national unadjusted mortality rate.” Should this be the national unadjusted mortality rate for COPD?

References

A

<http://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/hospital-value-based-purchasing/index.html?redirect=/hospital-value-based-purchasing>
<http://www.qualitynet.org/dcs/ContentServer?c=Page&pagename=QnetPublic%2FPage%2FQnetTier2&cid=1140537255912>

B

<http://www.qualitynet.org/dcs/ContentServer?c=Page&pagename=QnetPublic%2FPage%2FQnetTier2&cid=1140537255912>

C

KNG used 2009 100% Medicare inpatient claims data to identify Medicare beneficiaries admitted to short-term acute care hospitals with a principal diagnoses of one of the 3 initial conditions that will be included in the program (acute myocardial infarction, pneumonia and heart failure). To develop the analytic sample, we applied inclusion and exclusion criteria

consistent with the Medicare readmission measures endorsed by the National Quality Forum and used by CMS. Demographic and other characteristics of Medicare beneficiaries, such as age, sex, race, and dual eligible status were obtained from the 2009 Medicare denominator file. We were unable to use CMS's risk-adjustment methodology because 100% Medicare physician claims were not available to us. Instead, we used the Elixhauser comorbidity index for risk adjustment. The Elixhauser comorbidity measure is widely used in the literature as a risk-adjustment method for its proven predictive power of inpatient mortality. Following the approach used by CMS, we computed risk-standardized readmission rates (RSRRs) for each hospital and condition using a hierarchical regression model (HRM), which included hospital-level random effects.

**NQF Pulmonary Comment Period: AHA Letter Points for 30-Day COPD & Pneumonia Measures
(#0468, 0506, 1891, 1893)**

1. Failure to adjust for factors beyond the hospital's control.

These four measures, along with all of Yale's 30-day all cause risk standardized measures, consistently fail to adjust for factors beyond a hospital's control. Changes need to be made to these measures to properly stratify for patient characteristics (dual eligible status and race/ethnicity) and exclude extreme circumstances (transplant, end-stage renal disease, burn, trauma, psychosis and substance abuse). Further, these measures must address patient compliance with treatment plans and the quality of post-acute care provided to the patient.

CMS's risk-adjusted outcomes measures are developed with the patient perspective in mind. Recent studies demonstrate that hospitals can reduce the risk of readmission and mortality for their patients by optimizing care within the hospital and by facilitating patients' transitions to post-acute care. The measures adjust for demographic and clinical patient factors that affect outcome rates so that hospitals are not disadvantaged by their case mix, and so that differences in outcomes illuminate quality differences.

Consistent with NQF guidelines, the measures do not adjust for socioeconomic status (SES) or race. Any association between SES/race and health outcomes can be due, in part, to differences in the quality of health care received by groups of patients with varying SES or race. Risk-adjusting for patient SES/race would suggest that hospitals with low SES/minority patients are held to different standards for patient outcomes than hospitals treating higher SES patient populations. It could also mask important disparities and minimize incentives to improve outcomes for vulnerable populations. Again, the intention is for the measures to adjust for patient demographic and clinical characteristics while illuminating important quality differences.

Nevertheless, to inform consideration of the measure, we explored the potential implications of not adjusting for race or SES. Specifically, we assessed the extent to which hospitals' performance on the measures differed by the income levels of their patients or by the proportion of their patients who are African American. As shown in the NQF applications, for both COPD and Pneumonia we found little to no difference in the performance of hospitals with high proportions of low SES or minority patients compared with hospitals with low proportions of such patients.

AHA urges consideration of stratified reporting by patient SES; however, this approach would explicitly set different performance standards for different groups of patients. Given that our SES and race analyses show little to no difference in hospital performance by patient race and SES and given the downside of stratifying, there is not a compelling reason to stratify reporting for the measures.

We address clinical differences in case mix primarily by risk adjustment rather than through excluding patients with certain conditions from the measure. The goal in developing outcomes measures is to

create a clinically cohesive cohort that includes as many patients as possible admitted with the given condition. We aim to limit exclusions to only those factors that preclude fair assessment of care quality for an admission, such as lack of continuous enrollment, which prevents us from assessing patient risk factors, or patients leaving AMA, since hospitals do not have the opportunity to provide all recommend care for these patients. Greatly expanding our list of exclusions to include all the conditions mentioned in the comment would result in a measure that was less useful and meaningful, as it would reflect the care of the smaller number of a hospital's patients that presented without significant comorbidities. It also could create incentives for hospitals to code risk factors in order to exclude patients from the measures. To fairly profile hospitals' performance, it is critical to place hospitals on a level playing field and account for their differences in the patients that present for care. This is accomplished through adequate risk-adjustment for patients' clinical presentation rather than exclusion of patients.

For these measures we selected risk adjustment variables that were clinically expected to affect mortality and readmission rates and that met our statistical criteria. Using this approach, we selected some of the clinical factors AHA mentions, including 'renal failure' (CC 131; all measures), 'Drug and alcohol abuse' (CC 53; three of four measures), and 'other psychiatric disorders' (CC 60; three of four measures). Factors not included were not found to be critical to the model performance nor statistically significant.

Finally, AHA recommend that the measures address patient compliance and the quality of post-acute care. Although patients and other providers share responsibility for improving care outcomes, CMS expects hospitals to work to improve patient compliance and to arrange quality post acute care, and CMS believes it is therefore appropriate to hold hospitals accountable for the short-term outcomes of mortality and readmission without accounting directly for these factors. CMS is also continuing to develop quality measures, including readmission measures, for post-acute care providers.

2. All-cause readmission

The fact that these measures are “all-cause” means they will include readmissions and deaths that have nothing to do with the specific condition associated with the measure. A COPD patient could be readmitted/die from a fall, a car accident, or an unrelated condition not risk-adjusted in the measure. However, that readmission/death will still be included against the hospital in these measures. For these reasons, the hospital 30-day, all cause, risk-standardized measures² that are publicly reported provide information that is unsuitable for medical decision-making by consumers, because patients cannot parse out the many variables necessary to truly assess a hospital's performance.

As set forth in the technical report, we used all-cause readmission to assess performance, rather than readmission for acute exacerbations of COPD and for pneumonia, for several reasons. First, from the patient perspective, readmission for any reason is likely to be an undesirable outcome of care after an acute hospitalization. Second, readmissions not directly related to the COPD or pneumonia admission

may still be a result of the care received during the index hospitalization. For example, a patient hospitalized for COPD who develops a hospital-acquired infection may ultimately be readmitted for sepsis. It would be inappropriate to treat this readmission as unrelated to the care the patient received during the index hospitalization. Another patient might experience a hospitalization-related complication following the index COPD admission, which may go untreated and result in renal failure. The resulting readmission for renal failure could have been prevented with higher quality of care during the admission for COPD that reduced the risk for the complication. Furthermore, the range of potentially avoidable readmissions also includes those not directly related to the initial hospitalization, such as those resulting from poor communication at discharge or inadequate follow-up. As such, creating a comprehensive list of potential complications related to COPD hospitalizations would be arbitrary, incomplete and, ultimately, impossible to implement. Thus, the goal of this measure is not to reduce readmissions to zero, but to instead assess hospital performance relative to what is expected given the performance of other hospitals with similar case mixes. Finally, readmissions for rare reasons completely unrelated to hospital care, such as car accidents involving the patient as a passenger, are likely to be distributed randomly across hospitals and are not expected to introduce any bias into the measure results.

3. Reliability

AHA makes several remarks concerning “reliability”. This is often a confusing issue, because there are many different meanings and definitions of “reliability”; moreover, some reliability metrics refer to “intra-class correlation” (ICC), and there are several different metrics with this name as well. The AHA remarks mention three different reliability statistics, not all of which pertain to our measure. Our team is preparing a response to clarify these issues and will provide it to the Steering Committee next week.

4. Planned Readmission

The Affordable Care Act calls for these readmission measures to recognize planned cases and unrelated admissions; however these measures do not. [...] We urge the steering committee to take this into consideration and we ask the steering committee to query the measure developer on why this change has not been made to the measure. [...] We note that another NQF-endorsed measure (30-day all-cause all-condition risk standardized measure) does remove some planned readmissions. Therefore, pneumonia and COPD measures are not harmonized with the all-condition readmission measure. [...] We urge the steering committee to require the measure developer to harmonize these methodologies by also excluding these planned readmissions from the pneumonia and COPD readmission measures.

CMS recently developed an algorithm for identifying planned readmissions that is used in the hospital-wide measure and plans to adapt it for the COPD and pneumonia readmission measures. We will bring the updated algorithm and measure results back to the subsequent Steering Committee meeting.

COPD mortality measure

Based on comments in the “Evidence” section of the steering committee’s evaluation for this measure, it appears that the goal of this measure is to encourage hospitals to follow guidelines for COPD exacerbation management. However, the measure developer could create a more direct, process of care measure that evaluates whether the hospital followed such guidelines, or that measures their coordination of care, rather than creating a measure impacted by multiple outside factors.

We agree with the comment that process of care measures can be valuable for directing quality improvement efforts and assessing quality. However, they tend to measure only very narrow aspects of patient care and explain little of the variation in outcomes across institutions. In contrast, outcome measures assess a broad set of hospital and transitional healthcare activities that affect patients’ well-being. Patients who receive better, safer care during their hospital stays and during the transition to non-acute settings (e.g., home) will likely have improved outcomes such as survival, improved functional ability, and lower readmission rates. Ultimately, the outcomes are what matters most to patients. Our goal therefore is to develop and report scientifically sound outcomes measures such as the pneumonia and COPD mortality and readmission measures. The goal is consistent with CMS’ legislative mandates to expand the use of outcomes measures and with our National Quality Strategy.

5. Hospice enrollment

The exclusion of FFS Medicare patients enrolled in Hospice is a good attempt to delineate between patients going through aggressive treatment and those who have chosen to enter palliative care programs. It is not clear whether there are Medicare patients who are not part of this exclusion. In other words, why does the measure exclude FFS Medicare patients enrolled in Hospice as opposed to all Medicare patients enrolled in Hospice? In addition, this exclusion does not completely remove from the measure population all of those patients who are terminally ill, because Hospice programs may be full. It also does not appear to remove sicker patients who have signed a DNR measure. In short, the risk adjustment and exclusions do not seem to identify and address the sickest patients.

The measure exclusion covers all Medicare patients in the measure since the measure is limited to Medicare FFS patients (because data on Medicare Advantage patients is not available for the measure calculations).

The COPD and pneumonia mortality measures exclude patients who are enrolled in Medicare hospice on or before the day of admission because the goal of the hospitalization for these patients is likely not survival. However, consistent with guidelines for health care quality outcome measures, the 30-day measures do not exclude patients who transitioned to hospice or palliative care during their hospital stay because such transitions may be the result of quality failures that have led to poor clinical outcomes, thus, excluding these patients could mask quality problems. Moreover, the use of palliative care during a hospital stay is not necessarily an indication that a patient is no longer seeking life-sustaining measures. Palliative care is focused on providing patients relief of symptoms. It is increasingly used by patients who are not at the end of life and, therefore, should not be used to exclude patients from a mortality measure.

We do not exclude patients from the measure who are DNR because that status does not indicate the goals of the patient's stay; that is, many DNR patients have the goal of survival. Importantly, the measures adjust for a number of factors associated with the likelihood that patients are at the end of their lives, including protein-calorie malnutrition, metastatic cancer, dementia, and age, so that hospitals treating older, sicker patients can be fairly compared to hospitals with a healthier case-mix.

6. Harmonization: Risk Adjustment

The exclusions for [the COPD mortality] measure lack consistency with the exclusions for the 30-day all-cause risk-standardized mortality rate for pneumonia. The PN measure includes an exclusion for patients discharged alive on the day of admission or the following day who did not get transferred, because it is unlikely the patient had a significant pneumonia diagnosis. Why would the same kind of exclusion, such as lack of a significant COPD episode, not be part of the 30-day COPD measure?

The goal of the cohort exclusions generally is to ensure a clinically coherent cohort of patients that will be defined consistently across hospitals. For each condition, we considered whether setting a minimum length of stay would contribute to that goal, and this exclusion differs based on the clinical condition. The COPD Working Group and TEP did not think the exclusion was important in defining a consistent cohort.

7. Technical report feedback

The description of the adjustment/stratification for this measure included in the draft report also notes that "the RSMR is calculated as the ratio of the number of "predicted" to the number of "expected" deaths, multiplied by the national unadjusted mortality rate." Should this be the national unadjusted mortality rate for COPD?

Yes.

Memorandum

Subject: Response to the American Hospital Association June 5, 2012 comments to NQF on reliability of the CMS 30-day, all cause, risk-standardized COPD and pneumonia mortality and readmission measures

From: Yale New Haven Health Services Corporation/Center for Outcomes Research and Evaluation

Through: Lein Han, CMS

To: The National Quality Forum (NQF)

Date: June 19, 2012

Introduction

In its letter to NQF on the 30-day all cause pneumonia and COPD mortality and readmission measures, AHA makes several remarks concerning “reliability.” Reliability is often a confusing issue, because there are many different meanings and definitions of “reliability;” moreover, some reliability metrics refer to “intra-class correlation” (ICC), and there are at least two different metrics with this name as well. The AHA remarks mention three different reliability statistics, only one of which pertains to our measure, and as described below, the AHA’s description of how we are using reliability statistics is not accurate.

In our NQF applications, we report a fourth measure of reliability that “is the degree to which repeated measurements of the same entity agree with each other;” in particular, the degree to which the same hospitals when measured twice using different sets of patients from those hospitals produce similar measurements. This very specific reliability metric is a function of how consistently the measurement tool measures a latent quality of the measured entity, and is defined using “ICC(2,1)”, an intra-class correlation defined by Shrout and Fleiss.¹ Conventionally, interrater reliability is referred to a “moderate” standard of 0.4,² and we have used a similar standard in our application.

One major source of confusion in the AHA memo stems from the following distinction. It is possible for a measure to be highly reliable, though for measurement of certain (usually small) hospitals to be unreliable. The distinction is analogous to that between a survey instrument being highly reliable (test-retest reliability) but certain subjects (e.g., non-native speakers) having unreliable measurements. In

¹ Shrout P, Fleiss J. Intraclass correlations: uses in assessing rater reliability. *Psychological Bulletin*, 1979, 86, 420-3428.

² Landis, J. R. and Koch, G. G. (1977) "The measurement of observer agreement for categorical data" in *Biometrics*. Vol. 33, pp. 159-174.

order to better to clarify the distinctions and respond to the AHA memo, we make some brief definitions.

First, we define two ICCs:

ICC₀ : intraclass correlation, a parameter of a random effects model, equal here to the ratio of the between hospital variance to the total variance (between hospital plus within hospital (patient) level).

ICC(2,1) : intraclass correlation describing the test-retest reliability of multiple measurements of the same quantity; here, the measurements are readmission or mortality rates for a hospital, and the “clusters” are the n=2 measurements made on two groups of patients.

And then four reliability metrics:

R1 : test retest reliability: $R1 = ICC(2,1)$.

R2: hospital measurement reliability, used to inform decisions concerning minimum sample size for reporting; $R2 = n \cdot ICC_0 / (1 + (n-1) \cdot ICC_0)$.

R3: approximate signal to noise ratio; this metric is used in an Office of Clinical Standards and Quality (OCSQ) memo on reliability; it does not use either ICC₀ or ICC(2,1).

R4: metric for data validation of claims based measures, calculated as the percentage of measures that a hospital accurately reports. For example, assuming $R4=0.75$, then if a hospital reports 8 claims-based measures, chart re-abstraction should find 100% agreement on $\frac{3}{4} \times 8$ of those measures.

Of these four reliability metrics, **our NQF application refers only to R1**, although CMS has used R2 to inform decisions about the minimum hospital volume for public reporting. The AHA comments on reliability focus on concerns regarding minimum volume requirements. Given these definitions, we now discuss the specific AHA comments.

Response to Specific Comments

AHA states: *On February 13 CMS released a reliability study, required by the ACA, for claims-based measures. The study shows that the majority of claims-based measures currently used in CMS’s programs are unreliable. Among the measures included in the study was the pneumonia mortality measure.*

The memo (not study) referred to here is a CMS Office of Clinical Standards and Quality (OCSQ) memo on the reliability of measuring the quality of an individual hospital for measures in the outcome domain of the FY2014 Hospital Value-Based Purchasing Program. The memo focuses on informing decisions concerning the length of the data period and minimum N for the VBP

program. It uses the **metric R3** described above. It makes no statements about the overall reliability of the measures, only the reliability of measurements for hospitals with smaller sample sizes.

AHA states: *CMS's standard for non-claims-based measures is $R = 0.75$.*

The link that the AHA comment references to support this statement addresses standards for data validation. CMS sets a standard for data quality for hospital-reported (“non claims based”) measures. For these measures, charts are re-abstracted for a random sample of hospitals on a rolling basis, and the re-abstracted numerators compared to those reported originally. Hospitals reporting M measures must have 100% agreement on $\frac{3}{4}M$ (i.e., 75%) of the measures. This is the **metric R4** defined above. Though referred to as “reliability”, this has no relationship with either measure “reliability” (reported in our NQF application) or hospital measurement reliability (referred to in the previously cited memo) except in the shared nomenclature.

AHA states: *Further, we note that in order to achieve $R = 0.4$ for the pneumonia mortality measure, CMS recommends that a minimum of 211 cases are needed.*

Careful reading of the OCSQ memo referenced shows that CMS does not “recommend” that 211 cases are needed. The OCSQ memo concludes: *The results suggest that a similar length of time (24 months) is needed for CMS 30-day measures to achieve the same standard [of moderate reliability for a majority of hospitals].* CMS currently publicly reports the measure using 36 months of data, and the OCSQ memo does not contradict that approach.

AHA states: *The current minimum threshold for the Yale pneumonia measure is 25 cases—a major difference.*

This sentence refers to a sample size based on the results of **metric R2**, defined above. CMS has used this metric to inform decisions on the minimum sample size for reporting outcomes measures. This OCSQ memo uses an alternative method, metric R3, to explore the minimum sample size. As footnoted in the OCSQ memo: *“Using the ICC estimated from the regression parameters, higher values for reliability [R2] are obtained. When that method is used, over half of hospitals attain moderate reliability for heart failure and pneumonia with 12 months of data, and for AMI with 24 months.”*

AHA states: *Though the measure developer uses 36-months of data and this study does not address 36-months of data, we still do not believe this measure is reliable. **At a minimum, we urge the steering committee to request this reliability data of the measure developer for both the pneumonia and COPD mortality measures.***

If AHA is referring to test-retest reliability, we note in our application that using 18 months of data and metric R1 the measures meet the conventional standard. This is calculated using a split sample comparison. To calculate the same statistic for 36 months of data would require combining 6 years of data in order to have 36 months equivalent in each split sample. We feel that using 6 years could introduce temporal trends and other patterns that would reduce the overall consistency of the measurement sample.

If AHA is referring to minimum sample size for reporting, we note that the OCSQ memo supports a minimum of 24 months, compared to CMS' 36.

KNG, an independent research firm, has replicated CMS's reliability study for both the pneumonia and COPD readmission measures. The results of this analysis are included in Table 2, below. This data illustrates that both readmission measures fail to reach the industry standard of $R = 0.75$. KNG recommends that a minimum of 121 cases are needed for the pneumonia measure and 154 cases are need for the COPD measure in order to achieve $R = 0.4$.

It is not clear what reliability metric KNG used for this analysis; furthermore, KNG only used one year of data rather than the three years CMS uses for public reporting.

Summary

In our NQF applications we report a standard metric of test-retest reliability, R1, based on measuring each hospital using two exclusive sets of patients. The results for this metric indicate that the four outcomes measures are reliable (using the conventional standard). The AHA memo discusses several different notions and measures of 'reliability' rather than the one that we use. We have clarified the context for each one and have noted the appropriate interpretation.

Physician Consortium for Performance Improvement
Updated responses for Pulmonary and Critical Care Steering Committee

We would like to thank the Pulmonary and Critical Care Steering Committee members for their comments and recommendations on the PCPI Community-acquired Bacterial Pneumonia measures. We offer the following responses for these issues. A final determination has not yet been made on some of the issues but we continue working to find a suitable resolution.

Issues identified by Steering Committee

Measure (s)	Issue	Response
<ul style="list-style-type: none"> • 0096 Empiric antibiotic therapy for CAP • 0232 Vital signs for CAP • 1895 Mental status evaluation for CAP 	<p>Indicate the setting (outpatient/ambulatory care) in the title or description</p>	<p>All PCPI measure sets indicate the setting for which those particular measures are specified. Furthermore, the setting is supported by the codes and specifications included within a measure set. For NQF submissions, we strive to indicate the appropriate setting(s) in Sections 2a1.34-35 of the NQF submission forms.</p> <p>The setting for the Community-acquired Bacterial Pneumonia measures is the physician's office or emergency department.</p> <p>Accordingly, In the NQF submission forms, we indicated these settings as Ambulatory Care : Clinic/Urgent Care Ambulatory Care : Clinician Office Post Acute/Long Term Care Facility : Nursing Home/Skilled Nursing Facility Other: Emergency Department</p> <p>We regret that we may have confused the Steering Committee by also indicating the settings as Home Health Hospital/Acute Care Facility which suggested these were for the inpatient setting. This was an error on our part.</p>
<ul style="list-style-type: none"> • 0096 Empiric antibiotic therapy for CAP • 0232 Vital signs for CAP • 1895 Mental status evaluation for CAP 	<p>Why add "bacterial" to "community acquired bacterial pneumonia" when a specific organism is not identified in most patients?</p>	<p>The PCPI CAP Work Group uses the term "bacterial pneumonia" to clearly indicate that these measures apply to bacterial pneumonia since the standard of care for viral pneumonia is vastly different, especially as it pertains to antibiotic therapy. The Work Group did not want to propagate an assumption that the measures were meant to assess care of viral pneumonia for which a different set of measures would be most appropriate.</p>

0096 Empiric antibiotic therapy for CAP	Specify IDSA/ATS guidelines in the measure specifications	<p>The measure specifications as included in Section 2a1.3 of the NQF submission form indicate the use of an appropriate empiric antibiotic as defined by current ATS/IDSA guidelines. Furthermore, the ATS/IDSA guidelines are cited in Section 1c.17 of the NQF submission form as:</p> <p>IDSA/ATS 2007 guidelines Mandell LA, Wunderink RG, Anzueto A, et al. Infectious Diseases Society of America/American Thoracic Society Consensus guidelines on the management of community-acquired pneumonia in adults. Clin Infect Dis. 2007;44:S27-72.</p> <p>We will ask NQF staff for clarification regarding what materials are made available to members of the Steering Committee.</p>
0096 Empiric antibiotic therapy for CAP	Harmonization with hospital measure 0147 – use either “empiric” or “initial” but not both – confuses audiences; are other aspects of the measure harmonized? What efforts have been taken to harmonize the measures?	<p>PCPI staff is in contact with CMS staff. We continue to discuss harmonization issues between PCPI Measure 0096 and CMS Measure 0147*. However, we have not yet reached an agreement regarding harmonization. Furthermore, harmonization and language changes suggested for individual measures will require approval from our measure development panel, for which additional time will be needed</p> <p>We hope that the lack of a final determination on these measure-specific recommendations will not preclude the continued endorsement of the pneumonia measures.</p> <p>*CMS Measure 0147: Initial antibiotic selection for community-acquired pneumonia (CAP) in immunocompetent patients</p>
1895 Mental status for CAP	It is really “disorientation” or “confusion” and not “mental status”. Consider changing the wording.	<p>We cannot yet confirm any language change suggested by the Steering Committee until we have assured approval from our measure development panel, for which additional time will be needed.</p> <p>We hope that the lack of a final determination on these measure-specific recommendations will not preclude the continued endorsement of the pneumonia measures.</p>